Religious Divisions, Production Technology and Firm Productivity: Experimental Evidence from India

Arkadev Ghosh *

Preliminary: Please do not cite

June 8, 2021

Abstract

This paper shows that a firm's production technology mediates the effects of religious diversity on productivity. I randomly assigned Hindu and Muslim workers of a modern manufacturing plant in India to religiously mixed or homogeneous teams. Production tasks at the factory determine the nature of contact between co-workers. I categorize tasks as High or Low Dependency based on the degree of continuous coordination required for production and the dependence on co-workers for breaks. Mixed teams are less productive than homogeneous teams in High Dependency tasks, but diversity does not affect productivity in Low Dependency tasks. Religious diversity in High Dependency tasks leads to greater free-riding as well as more inter-group blame. The negative effects on productivity however attenuate over time, subsiding completely in 4 months. Overall, while productivity effects of religious diversity are more negative for High Dependency tasks, the effects on social preferences are more positive. This pattern of results suggests that there is a tension between the goals of maximizing productivity versus improving inter-group relations.

*VSE, University of British Columbia, (email: arkadev@mail.ubc.ca). I thank Siwan Anderson, Munir Squires, Matt Lowe, Patrick Francois, and Jamie McCasland for their advice, support and guidance throughout the project. I would also like to thank Ashok Kotwal, Claudio Ferraz, Felipe Valencia Caicedo, Thorsten Rogall, Thomas Lemieux and Anand Chopra for useful suggestions and comments. Financial assistance from CIDER is gratefully acknowledged. This project received ethics approval from UBC BREB, and was pre-registered in the AEA registry with ID #0004270.

1 Introduction

There is an extensive literature suggesting that cultural diversity leads to poor economic outcomes.[1] Much of this work is focused on coordination failure over public good provision (Easterly and Levine, 1997; Alesina and Spolaore, 1997; Miguel, 2004), but there is also evidence of how poor social ties (Lazear, 1998; Afridi et al., 2020) and taste-based discrimination (Becker, 1957; Hjort, 2014) can lead to low performance in the private sector. This paper contributes to this literature by exploring whether a firm's production technology can mediate the effects of diversity on productivity.

I implemented a unique experiment in a modern factory in West Bengal, India that employs both Hindus and Muslims. The nature of contact between workers is determined by the manufacturing task performed. I categorize team tasks into two pre-registered types: High Dependency (HD) and Low Dependency (LD). This classification is based on the degree of continuous coordination required amongst workers performing a task to ensure uninterrupted production, and the dependence on teammates for breaks. Worker efforts have a high degree of complementarity in HD tasks while in LD tasks they are typically non-complements.[2]

I randomly assigned 586 workers to religiously mixed or Hindu-only teams. I could not further randomize workers to tasks as the firm's policy does not allow workers to sort into jobs during recruitment or over their tenure. But the randomization design together with this policy feature allows me to attribute potentially different effects of religious diversity in HD and LD tasks to production function differences, as opposed to differences in worker types in these

[1] It is possible that diversity could lead to positive economic outcomes due to strategic complementarities in interacting with outgroup individuals (Artiles, 2020; Montalvo and Reynal-Querol, 2017) or under certain requirements of ethnic interaction (Bhalotra et al., 2018; Marx et al., 2021).

[2] An example of a HD task is work on a fast moving conveyor belt where each worker is responsible for collecting every 2nd or 3rd piece of a product on the belt. If one worker cannot keep up with the pace of collection, the machine speed needs to be reduced affecting the productivity of all workers. Oven tray washing is example of a LD task, where each worker is responsible for washing a certain number of trays every hour – the productivity of one worker does not directly or immediately influence other workers. Detailed description of types of tasks as well as how they are classified into HD and LD follow in section 2.

jobs. The experiment was designed to estimate the effects of diversity (by task type), using overall line-level output (comprising of a series of tasks) as well as individual task level performance. I kept the randomized teams intact for a period of 4 months in order to estimate dynamic effects of intergroup contact on both productivity and preferences.

This first key finding of the study is that religious diversity negatively affects team output in HD tasks but not in LD tasks. Mixed teams in HD tasks have 5% lower output than Hindu-only teams. This effect is economically significant – going from having all mixed teams in HD tasks to Hindu-only teams, would increase daily output value by Rs. 675,000 ($9300) at the firm. By contrast, I find no output loss from having religiously mixed teams in LD tasks. The second key finding is that the difference in output between mixed and non-mixed HD teams attenuated significantly over the treatment period. From a difference of 20% in output at the beginning of the experiment, the effect reduced to less than 2% by the end of the 4th month.

Surveys of the workers reveal certain mechanisms at play to explain these findings. Regarding the first finding, workers in HD mixed teams complain 3.8 percentage points (29%) more about free-riding by co-workers, report 3.7 percentage points (46%) greater inter-group blame and mention lower team cohesion relative to homogeneous teams in general. Consistent with the null effect on output, workers in mixed LD teams are less likely to complain about their co-workers relative to those in mixed HD teams. Regarding the second finding, comparing baseline and endline survey responses, Hindu workers in mixed HD teams report that they are more comfortable communicating (40%) and taking orders at work (25%) from Muslims, and are also more willing to be put in a mixed team (43%) in the future, relative to those in homogeneous teams. This is consistent with the attenuating output effects in HD tasks.

The null effect of diversity on productivity in LD sections helps to rule out certain mechanisms. Due to social reputation concerns around in-group members, Hindu workers in homogeneous LD teams could exert greater effort than those in mixed LD teams making Hindu-only LD teams more productive (Afridi et al., 2020). Workers could simply sabotage outgroup mem-

bers due to taste discrimination (Hjort, 2014). It is important to note that even though worker efforts are typically non-complements in these tasks, teams are still required to coordinate on many aspects of production such as sharing raw materials, following up on teammate's progress to plan one's own schedule of work, collaborating during production breakdowns etc., whereby it possible to undermine the efforts of teammates. However, I do not find these effects to be present.[3] This suggests that the negative effects in HD sections are caused by the frictions that arise when working in tight-knit production environments with non-coreligionists, as opposed to workers simply sabotaging outgroup colleagues.

Production technology may mediate effects of diversity on productivity, but does production technology also mediate effects on intergroup relations? The former is the relevant question for a firm, in deciding whether to integrate workers and the latter is important for a state aiming to reduce intergroup conflict. While there is no output loss from religious diversity in LD tasks, there are little or no positive effects on attitudes of Hindu workers, unlike in HD tasks.

This pattern of results suggests that there is a tension between the goals of maximizing productivity versus improving intergroup relations. More speculatively, the results might help explain why in equilibrium there can be a lot of integration at work without intergroup relations improving – the integration might only occur in contexts where intergroup contact is socially ineffective. Certain negative aspects of production in developing countries, such as the presence of too many small firms (Hsieh and Olken, 2014) as well the absence of large, complex assembly-line production processes (Kremer, 1993), might, at least in part, be attributable to the difficulty of organizing High Dependency production with ethnically diverse workforces.

I present a theoretical model to rationalize these core findings, especially the attenuation in output differences between mixed and non-mixed HD teams over time. The model makes

[3] During a brief period of religious tensions in West Bengal following the passing of the Citizenship Amendment Act (CAA) in India, which coincided with the intervention period, I find religious diversity to have negative effects in LD tasks too. This effect is short-lived. Nevertheless, this shows that LD tasks are not a placebo type where frictions among workers do not matter at all. Rather the loose-knit production structure mean that output in these tasks is less sensitive to such frictions – but extreme events can lead to workers sabotaging out-group members.

additional predictions with respect to treatment heterogeneity by worker characteristics which are consistent with the data. In my framework, the minority group (Muslims) initiates the integration process with the majority group (Hindus), which gradually ameliorates productivity losses. They exert higher effort than optimal (in the short run) to convince the majority group (Hindus) that Muslims are not lower productive individuals.[4]

Consistent with majority-minority relations, in factories and even other formal workplaces, Muslims are generally used to working alongside Hindus, while a large share of Hindus are not used to working with Muslims.[5] Given this background, I assume that on average Hindus have lower priors regarding how hardworking their Muslim co-workers are, relative to in-group Hindu co-workers.[6] Muslim workers do not make this distinction. When randomized into a mixed HD team, where team success is determined by the joint effort of all workers, Hindus optimally choose low effort based on their initial (low) prior about the effort level of their Muslim co-workers. This leads to lower output in mixed HD teams initially. Team output is a noisy measure of worker effort in this framework. Hindu workers update their beliefs about Muslims based on their own effort and realized output. Muslim workers, who as explained, have typically had much greater contact with Hindus are forward looking and internalize this behaviour of Hindu workers when choosing their optimal *effort investment path*. Given a long time horizon, it is optimal for Muslim workers to exert high effort despite the fact that Hindus initially exert low effort. This is because Muslim workers can persuade Hindu workers to eventually exert high effort as the latter begin to observe *greater realizations of high output days*, than ex-

[4]The model treats the majority (Hindus) and minority (Muslims) groups differently. Hindu-Muslim conflict has been a recurring phenomenon in India since partition and independence in 1947, with Muslims suffering greater discrimination and violence against them, as well as bearing larger economic losses (Mitra and Ray, 2014). Muslims in India continue to lag behind Hindus on various economic indicators including income and education (Asher et al., 2018), face social exclusion (Alam, 2010) as well as discrimination in the labor market (Kalpagam et al., 2010; Khan, 2019) due to their minority status.

[5]I show evidence of this in my data. Roughly 50% of the Hindu workers in the firm worked in homogeneous teams at baseline, while all Muslim workers worked alongside Hindus. Similarly, 43% of Hindus reported to have no contact with Muslims outside of work, whereas only 9% of Muslims reported the same about Hindus. This is due to the respective shares of Hindu and Muslims workers in the factory and in general in the population in India.

[6]I show evidence that Hindus and Muslims are equally productive as individuals.

pected under low effort from their Muslim teammates. The incentive for Muslim workers to bear the “integration cost" is that a high steady state output is beneficial in the long-run for career progression in the factory. In LD tasks, worker efforts are non-complements whereby the effort levels of Hindu workers are not dependent on their priors about Muslims. As a result, team output is not affected by diversity. Consequently, Muslim workers do not need to persuade Hindus to update their priors.

I explicitly study interactions between Hindu and Muslim workers during production and find the data to be consistent with the mechanism in the model. A threat to identification is that there are no Muslim-only teams in this study. If Muslims have lower productivity at HD tasks[7], the treatment effects on output would simply be picking up differences in average productivity between mixed and Hindu-only (homogeneous) teams. This is however unlikely for three reasons. First, I show that the negative effect on output in mixed HD teams is significantly smaller, if Hindus in the team have had Muslim co-workers in the past. This is consistent with the mechanism in the model. Second, the large attenuation in the treatment effect over time is unlikely if Muslims were simply unproductive at HD tasks. Finally, I show that there is heterogeneity in this attenuation; teams in which Hindus have had greater contact with Muslims suffer smaller losses initially relative to teams in which Hindus have had no contact. The effects completely dissipate for the former group, but remain negative and significant for the latter by the end of the intervention. These dynamics are hard to explain through a framework in which lower output in HD mixed teams is caused simply by lower productivity of Muslim workers.

This paper contributes to a few different sets of literature, the most relevant of which is on ethnic diversity and firm production. The evidence is scarce with a few papers documenting negative effects on productivity (Hjort, 2014; Afridi et al., 2020; Parrotta et al., 2012; Hamilton et al., 2012; Churchill et al., 2017). Hjort (2014) exploits quasi-random variation in the ethnic composition of teams in a Kenyan flower plant and finds that ethnically mixed teams have lower

[7] This is of course not true in LD, because diversity does not affect productivity in LD tasks.

productivity due to taste-discrimination. Afridi et al. (2020) exploit variation in caste composition of teams caused by worker absenteeism in Indian garment factories and show that caste homogeneity boosts productivity. These papers typically look at similar tasks performed by all the workers in a factory. Mine is the first paper to show how differences in the nature of contact amongst workers driven by production technology differences, affects team productivity. A second key feature of my research design is that I estimate dynamic effects of repeated inter-group contact on team production and social preferences in the same setting. Studies that exploit frequent team switching in observational data from firms (for identification), are not able to identify such effects. My results emphasize the need for intergroup contact to occur for a sufficiently long period of time. Diversity leads to a *low output equilibrium* in my model when contact is only short-lived, since workers do not have the incentive or the time to update their priors sufficiently about outgroup colleagues. This could possibly explain why, in settings where teams are frequently switched, researchers do not find that a history of being in teams with outgroup members leads to reduced prejudice and discrimination (Hjort, 2014).

The next set of papers my work adds to are on social preferences at the workplace (Bandiera et al., 2010, 2013; Mas and Moretti, 2009; Carpenter and Seki, 2011; Hjort, 2014; Ashraf and Bandiera, 2018). I show that in the Indian context, factory workers discriminate against non-coreligionists leading to loss in the firm output. The plant I study offers a flat monthly wage to its employees. The wage level is based on seniority and experience at the firm. This is different from the setting in majority of other papers on this topic, which study team productivity under group versus individual pay structures. I show that even without explicit pay incentives (based on overall team productivity or individual productivity), social preferences at the workplace can have large effects on team productivity.

This paper also relates closely to the literature on how social preferences are formed (Fershtman and Gneezy, 2001; Jakiela et al., 2011; Rao, 2019) and how the effects of inter-group contact depend upon the *type and nature* of contact (Allport et al., 1954; Pettigrew et al., 2011;

Paluck et al., 2019). The majority of existing empirical studies largely analyze one type of contact (Burns et al., 2015; Carrell et al., 2015; Schindler and Westcott, 2015; Broockman and Kalla, 2016; Mousa, 2018) or use non-randomized variation in the type of contact (Pettigrew et al., 2011; Bazzi et al., 2017). Lowe (2020) is the first to show experimentally that intergroup contact has different effects based on the type of contact (collaborative or adversarial) by randomly allocating men to mixed or homogeneous caste cricket teams in India. My paper is the first to show that the nature of contact determined by the degree of complementarity in worker efforts matters for economic gains in a real-world team production setting.

Finally, the literature on employer learning in the U.S. (Farber and Gibbons, 1996; Altonji and Pierret, 1998; Altonji and Pierret, 2001) argues that if firms discriminate amongst workers based on easily observable characteristics (education and race for example), then as employers begin to observe (noisy) indicators of workers' performances and learn about their productivity, the initial information should gradually become redundant. This paper shows that such updating also holds true for incorrect priors that individuals might have about their co-workers (not just employees) in a team production setting.

The rest of the paper is organized in the following manner. Section 2 describes the firm that I study - specifically the layout of production lines and sections, how shifts operate, the pay structure of workers and the different tasks involved in the production of the final items. It also describes how tasks are categorized into High and Low Dependency types in more detail, elaborating on key characteristics of each task and what they mean for the nature of contact amongst workers. I discuss randomization details and balance checks in section 3. Section 4 presents the econometric specifications used to estimate treatment effects using both production and survey data. The results are presented in section 5. In section 6, I present a theoretical model to rationalize the main results and confirm additional patterns in the data that are consistent with implications of the model. Alternative explanations are considered in section 7. In section 8, I present results from a survey of production supervisors of multiple firms, designed primarily

to infer their understanding of how diversity affects productivity and steps they are willing to undertake to mitigate possible negative effects. Finally, section 9 concludes.

2 Production at the factory

This section describes the production lines and sections at the factory and illustrates how shifts operate. It discusses the pay structure of workers and also reports characteristics of the workers by religion at baseline.

2.1 Production lines and worker characteristics

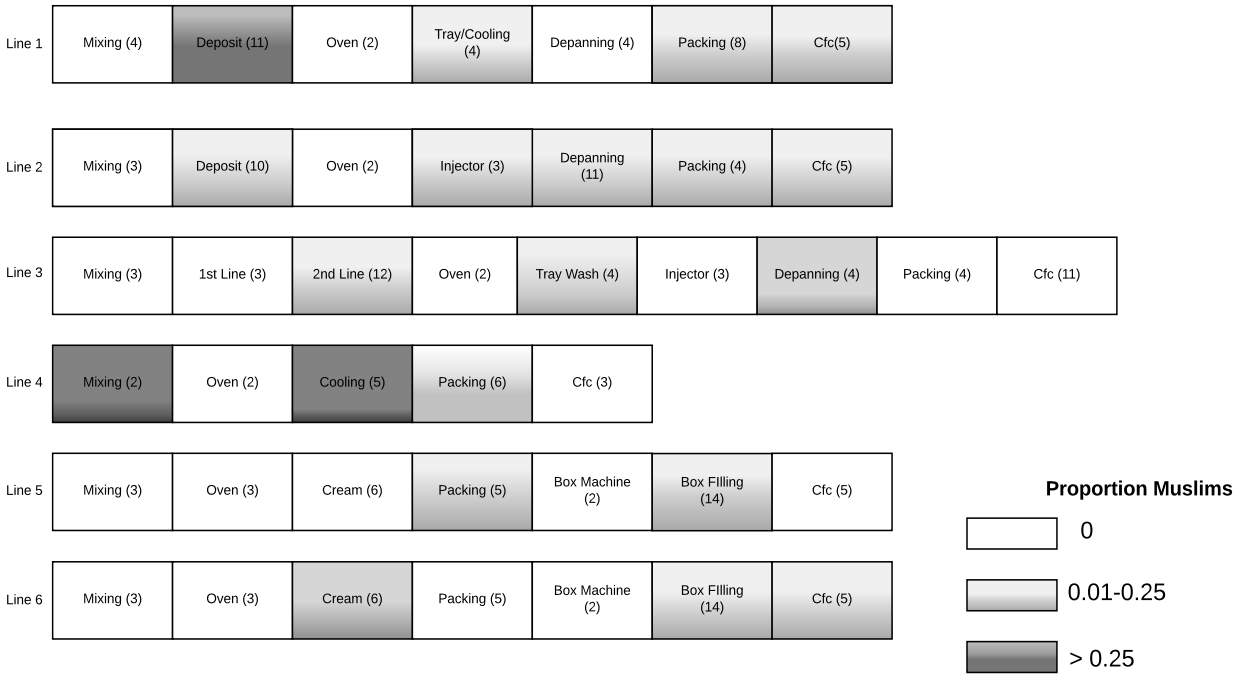
Production lines, sections and shifts

Production at the factory occurs in three different shifts - morning, afternoon and night. There are 3 cohorts per production line, who as a team rotate shifts on a weekly basis.[8] As a result, workers have fixed teams both at the line-level and line-section-level, i.e. their co-workers do not typically change, only their shift of work as a team changes from one week to the next. Figure 1 shows the structure of production lines at the factory. There are 6 production lines in total. Each line is sub-divided into sections (small blocks in the figure) based on the production task that is undertaken. [9] The numbers in parenthesis denote the count of workers in each of these sections. The overall structure of each line is similar; raw materials are first mixed, then baked and finally finished products are packed. However, each of these broad tasks could occur over multiple sections. Some sections have operators - these are individuals who operate machines and are typically senior to other workers.

[8] Teams move from morning to night to afternoon shifts.

[9] Occasionally workers are moved across shifts and lines. This is determined by worker absenteeism and turnover.

Figure 1: Structure of Production Lines



Note: This figure shows the structure of all 6 production lines in the factory. The numbers in parentheses denote the count of workers in each section per cohort. Each production line has 3 cohorts working on it in each of the 3 shifts in a day. The color shades denote the proportion of Muslims workers in each section in one particular cohort at baseline. The composition was similar in the other cohorts as well. Please refer to figure D.4 in the Appendix for this figure without the color shades.

Religious composition of production lines

Table 1 shows the religious composition of all cohorts for each production line at baseline. Line 4 only has two cohorts (morning and afternoons shifts only) while all other lines have three teams or cohorts each.[10] While there is variation in the proportion of Muslims across teams, it is clear from this table that Hindus and Muslims are not segregated in particular lines or shifts at the factory. On average, each line and cohort roughly has between 15%-20% Muslim workers, which is very close to the overall share of Muslims in the factory. This is formally shown in Figure D.6 (Appendix) – I regress a dummy variable denoting a worker's religion on line and cohort fixed effects and show that balance in religious composition across production lines and/or

[10] I use the words team and cohorts interchangeably.

cohorts cannot be rejected.

Table 1: Proportion Muslims at baseline by line-level team and cohort

Line	Cohort 1	Cohort 2	Cohort 3	Average
Line 1	0.26	0.18	0.23	0.22
Line 2	0.20	0.11	0.22	0.18
Line 3	0.11	0.27	0.08	0.15
Line 4	0.22	0.28	-	0.25
Line 5	0.12	0.07	0.19	0.13
Line 6	0.14	0.20	0.07	0.14
Average	0.17	0.19	0.16	0.18

The fact that Muslims are a minority in the factory together with the structure of production lines that require small worker teams even within lines, means that a large section of Hindu workers have little or no contact with their Muslim counterparts. This can be observed in Figure 1, where the religious composition of production sections for all 6 lines, in one particular cohort is shown. A large number of sections (close to 50%) have no Muslim workers at all. The share of Muslim workers in majority of the other sections is less than 0.25, with only two sections having a greater proportion than that (shaded dark grey). This is similar across the other two cohorts as well. This is important for two reasons. First, this tells us that a large fraction of Hindu workers typically have (had) no contact with Muslim workers – therefore one would expect the degree of inter-religious contact induced by the treatment (60% Hindus and 40% Muslims in mixed teams) to impact productivity and preferences on average. Second, there is some variation in inter-religious contact at baseline that will allow heterogeneous analysis of treatment effects. This can be used to test predictions from the theoretical model (Section 6) that rely on existing priors of Hindus about Muslims, which is likely to be determined by the history of past contact.

Pay structure of workers

Workers at the factory are paid a flat monthly wage based on their experience (tenure) and

level of expertise on the job (skill). Wages are therefore not dependent on individual daily productivity; however poor performance over a long period of time can lead to workers being moved to a lower skill group which would affect salary. Workers are categorized into unskilled, semi-skilled and operator groups. Approximately 80% of workers are unskilled and the rest are semi-skilled or operators. Semi-skilled workers undertake the same tasks as unskilled workers, while operators are in charge of handling machines.

Characteristics of Hindu and Muslim workers

Summary statistics of worker characteristics by religion are reported in Table C.8 (Appendix). It can be observed that workers are not sorted into HD and LD jobs differentially by religion. There are however important differences between Hindus and Muslims. Muslim workers have lower schooling and tenure at the factory, relative to Hindu workers. It has been documented in other studies as well that Muslims on average tend to have lower education relative to Hindus in India (Bhaumik and Chakrabarty, 2009). The difference in average tenure however might come as a surprise. In the district where the factory is located, Muslims have traditionally been tailors, which many families still continue to pursue as their business. As a result, this network allows Muslims to work in this sector, providing them with an outside option of employment. This is often cited by the management as a factor behind the larger turnover of Muslim workers.

Muslim workers report having much greater contact with Hindus outside of work, which is expected given that Hindus form the majority group in the area and across India in general. Consistent with this, Muslims report to be more comfortable than Hindus when it comes to communicating with non-coreligionists. Surprisingly, both groups report to be equally uncomfortable taking orders from non-coreligionists at work. Finally, as shown in Table C.8, there are also differences in political preferences between Hindus and Muslims as proxied for by support for the controversial National Registrar of Citizens (NRC), a bill which is thought to discriminate against Muslims.

2.2 Direct Dependency as a measure of production technology

Direct Dependency refers to the degree of continuous coordination required amongst workers performing a task to ensure uninterrupted production. I study it as the key aspect of production technology for two main reasons. First, a key distinction between HD and LD tasks relates to a core idea in economics: whether labor inputs are complements or not. Worker efforts have a high degree of complementary in HD sections while they are typically non-complements in LD tasks. Second, the degree of complementarity in labor inputs affect incentives to interact, suggesting that this might matter for the effects of religious divisions. Some key characteristics of High and Low Dependency sections are listed in Table 2.

Table 2: Characteristics of High and Low Dependency Tasks

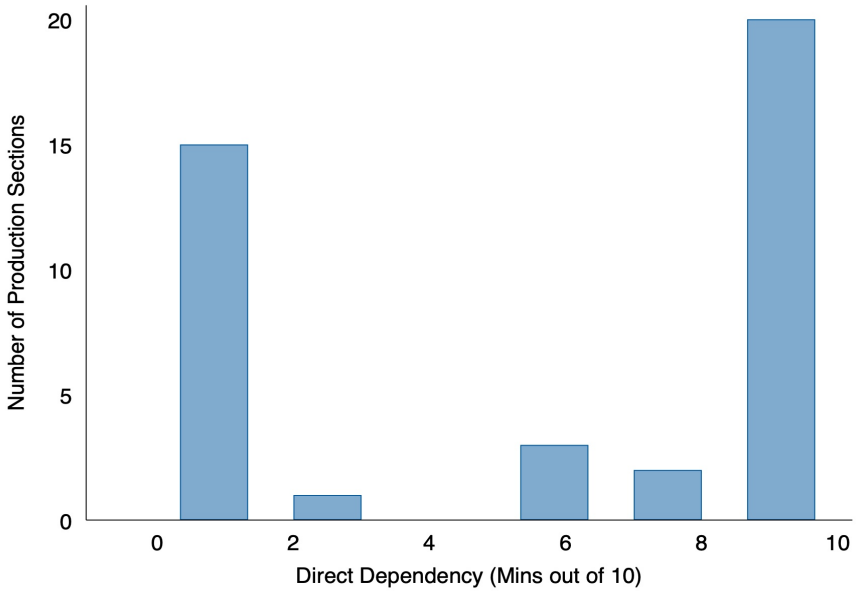
Work condition	**High Dependency (HD)**	**Low Dependency (LD)**
Task coordination	High and Continuous	Low and Intermittent
Control over breaks	Low	High
Physical mobility	Restricted	Good
Repetitive monotony	High (Machine Speed)	Low (Occasionally paced by machine)

Task coordination

The first key distinction between High and Low Dependency sections is the amount of continuous coordination required amongst co-workers. A high degree of continuous coordination is required in HD sections, whereas it is only intermittent in LD ones. I quantify this with time-use data. Research Assistants recorded minutes (out of 10) of continuous coordination required amongst workers for production to continue without interruption in each section. HD sections typically require workers to coordinate continuously for 9-10 minutes (out of 10), whereas the average in LD sections is only 2 minutes. Sections above the median value (≥ 9) on this scale are classified as HD sections and the rest as LD sections.

The distribution of Direct Dependency is shown in Figure 2. Most tasks require either high continuous interaction (9-10 minutes out of 10) or less than 2 minutes of continuous coordination - this leads to a bi-modal distribution as can be observed in the figure. This allows easy classification of tasks into one of these two groups - an important (third) reason to pick this measure as a proxy for the nature of contact amongst workers driven by production technology.

Figure 2: Distribution of Direct Dependency



Note: This figure shows the distribution of Direct Dependency. Enumerators visited all sections of all production lines and took stopwatch measures of the number of minutes (out of 10) for which workers were continuously dependent on each other for production to occur. The figure is generated from these stopwatch records.

Control over breaks/relief time

The second key distinction between HD and LD tasks is with respect to control over breaks during the production process. Due to the tight-knit structure of the work environment and dependence on co-workers every minute of the production process in HD tasks, each worker individually has little control over when they can take a break. By contrast in LD sections each worker has much greater control over scheduling breaks. Sub-groups of workers need to provide relief to other workers in the same HD section, a concept known as “relief time". There

are often arguments amongst workers regarding how to schedule these and when some workers take more time than allocated to them. Supervisors report that such disruptions are one of most common causes of production stoppages leading to lower productivity.

Physical mobility

Physical mobility is restricted in HD sections. For example, workers are typically required to stand close to each other on conveyor belts and pick products up as they move on the belt. Coordination with others doing the same is therefore also key. In LD sections, the nature of production and greater individual control over the process allows workers greater physical mobility.

Repetitive monotony

Repetitive monotony is higher in HD sections compared to LD sections. Work cycles are shorter, more frequent and typically more stressful as a result. The machine speed set by the supervisor often determines the speed of work, allowing workers little control over the process. If workers do not perform up to the mark, supervisors need to reduce machine speed which causes loss in output. Informal interviews with the supervisors made it clear that it is not uncommon for them to vary machine speed in these areas. This could happen due to worker absenteeism leading to changes in teams, as well as due to workers simply not coordinating as expected on certain days of production. By contrast in LD sections, workers typically have more control over process speed, and can re-allocate their time across different sub-tasks to a greater extent.

Direct Dependency and other section-level characteristics

In Figure D.5 (Appendix D), I show all 6 production lines at the production factory broken down into High and Low Dependency sections. A couple of points are worth mentioning here. First, some sections (tasks) in all lines are consistently categorized to be either HD or LD. For example, Mixing is always LD while CFC (Carton Packaging) is always HD. At the same time, some other tasks such as Packing could be HD in some lines while LD in others. This difference

is caused by the types of machines used, which in turn determines the level of dependency amongst workers.

In Table C.7 (Appendix C), summary statistics of various aspects of the physical environment of HD and LD sections are presented. I focus on factors which could act as potential confounders to the main mechanism in this paper. I measured the degree of non-work interaction (time workers spend chatting) and noise levels in each section of each production line and rule out that HD and LD sections are systematically different on these aspects of the physical work environment. The only statistically significant difference between HD and LD sections is in the average temperature; HD sections tend to be warmer by 2 degree Celsius. This difference is primarily due to a few colder LD sections in one particular production line. One could worry that hotter temperatures might intensify the negative effects of religious divisions, driving part of the effects that I find. This is not the case - all my results are robust to dropping this production line/sections from the analysis.[11]

3 Research design

I first go through the randomization process in detail and then present balance checks over a range of worker characteristics across different treatment arms.

3.1 Treatment and Randomization

As mentioned earlier, the factory operates in 3 shifts (morning, afternoon, night) and an entire cohort of workers move from one shift to the next on a weekly basis. A new set of workers come to work in each shift on a particular day. Therefore each line has 3 different teams (cohorts) working on it each day of the week. For the purpose of randomization, I did not move workers across production lines, but did so across cohorts.

[11] These results are available upon request.

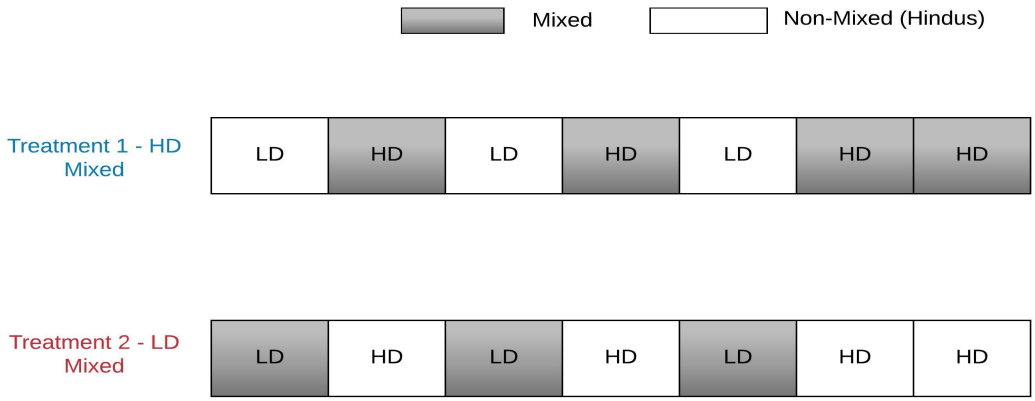
Workers were randomized into line-section-level teams such that overall there were two distinct types of teams (treatments) at the line-level in each production line. This is described in Table 3. The first type comprised of line-level teams that had all their HD sections mixed (HD-Mixed line), while the second type had all their LD sections mixed (LD-Mixed line). Therefore, two of the randomized cohorts in each line were of one team type while the third cohort was of the other type. Figure 3 below provides a visual illustration of the two types of line-level teams. I use Line 3 from Figure 1 for this illustration. Individual section names are replaced by HD and LD labels to denote section type. The first type of line-level team has all its HD sections mixed (shaded in grey) while its LD sections are comprised of only Hindu workers. The structure in the second type is exactly the opposite - LD sections have religiously mixed teams while HD sections have only Hindu workers (non-shaded). This leads to four different types of line-section-level teams: 1. HD Mixed 2. HD Non-Mixed 3. LD Mixed and 4. LD Non-Mixed. Whether a production line had 2 cohorts of HD-Mixed lines (and one of LD-Mixed) or the other way round was determined by the overall number of Hindus and Muslims in the line at baseline.

Table 3: Randomized Team Structure

Team Types (Lines)	**High Dependency Sections**	**Low Dependency Sections**	**Treatment**
1	Mixed	Non-Mixed (Hindus)	Treatment 1
2	Non-Mixed (Hindus)	Mixed	Treatment 2

The supervisors at the factory gave a daily rating to each line-section-level team based on their performance, independent of the performance of the overall line or other sections. I use this data to explicitly test for the extent of output loss caused by religious diversity in HD sections relative to LD sections.

Figure 3: Randomized Team Structure



Note: This figure shows the structure of line-level teams after randomization. HD Mixed (LD Non-mixed) lines or Type 1 had all their HD sections mixed and LD sections non-mixed. The opposite is true for LD Mixed (HD Non-mixed) lines.

I now describe the randomization process. Randomization was constrained by one key limitation - the number of workers switching their section of work (their task) had to be minimized. Even though the induction of workers to specific tasks (unless as an operator) takes only between 1-2 days, it is impossible to train all workers in new tasks simultaneously – this would lead to substantial interruptions and breakdown in production. The management was not willing to do this. As a result, the randomization process was designed such that did it not require majority of workers to change their section of work and hence the *Dependency* of their task at baseline. I address concerns with respect to selection of workers into HD and LD jobs subsequently.

The first step in the randomization process involved determining the number of Hindus and Muslims in each section of each production line. Since workers were not moved across production lines, this was typically constrained by the overall number of Hindus and Muslims in a line across the three cohorts at baseline. The religious composition of each production line at baseline was very close to the religious composition in the factory overall. Hence, mixed sections (both HD and LD) of all six lines ended up comprising of roughly 40% Muslim workers

after randomization.[12]

The second step in the process involved sorting workers by section × religion × skill (across all cohorts) and shifting workers across sections (tasks) in order to ensure that each section had enough Muslim workers (across all cohorts) required for randomization, as determined in the previous step. This is important because at baseline, not all sections of all lines had enough Muslim workers (sometimes none) such that the desired team structure in Figure 3 could be achieved. For example, the *Injector* section (task) in Line 3 had no Muslim workers at all across all three cohorts. In such cases some Hindu workers in that section would be randomly shifted out and replaced with randomly chosen Muslim workers from another similar section with enough Muslims. This process meant that at the end of step 2, all sections of all lines had both Hindu and Muslim workers[13] who would then be allocated to line-level teams. This also satisfied the management's requirement of minimum task-shifting.

Lastly in the third and final step, workers were sorted by their new section × religion × skill (worker or operator) and randomly allocated into line-section-level teams in order to achieve the line-level team structures, shown in Figure 3. Line-level teams were then finally allocated to each of the three shifts. A detailed description of each step involved in the randomization process is presented in Appendix A. Figure A.1, provides a visual illustration of the same, especially focusing on how section-shifting allows formation of the desired line-level team structures.

3.2 Data collection, timeline and attrition

Figure A.2 (Appendix), presents the timeline of the intervention and the sample size by treatment arms. The baseline survey was conducted in July and August 2019, the intervention started

[12] Note that the religious composition of a particular section in a line would be exactly the same across all shifts if they were the same line-level team type. In other words, if cohorts A and B in Line 1 were such that all their HD sections were mixed and LD sections were non-mixed, then each of their HD sections would have exactly the same ratio of Hindu to Muslim workers i.e. *Packing* in cohort A would have exactly the same number of Hindus and Muslims as *Packing* in cohort B. Non-mixed teams of course only have Hindu workers.

[13] This is required because for each section of each line there would at least be one line-level team where that section had to have a mixed team.

in November and ended in March 2020 when COVID-19 related restrictions were imposed in India. Note that the number of workers interviewed at baseline is larger than those part of the actual experiment. This is because the factory decided to lay off some workers due to low product demand in two lines (Lines 5 and 6) after the baseline survey, but before the randomization began. A total of 586 workers were part of the intervention of which 546 could be reached at endline for the phone survey (attrition rate 6.8%). An online survey was conducted with more than 100 production supervisors across 5 different firms in April 2021. This survey was mainly focused on inferring whether supervisors understand the effects of diversity on production, and if so, what policies do they already have in place or are willing to take in the future to mitigate possible negative effects. The results from this survey are presented in section 8.

3.3 Randomization check

I now proceed to balance checks presented in Table 4. Outcomes are divided into two broad categories - (1) those that are relevant at work (Panel A) and (2) other outcomes (Panel B). The unit of analysis here is an individual. The main regressors are the interaction terms Mixed × LD and Mixed × HD which denote the type of line-section-level team and hence the treatment status of an individual. Line × Section fixed effects are included in these specifications, whereby the main effect of HD versus LD is not separately identified. The omitted group is therefore all workers assigned to non-mixed teams. We can see that across a wide range of characteristics, which include factors that are relevant at the workplace (such as tenure, age and past contact with non-coreligionists) as well as other general attributes (such trust, altruism and contact outside work), workers are similar across the treatment arms. I use this particular specification for balance checks because the same specification is used to estimate treatment effects at the line-section-level on team production as well as individual survey outcomes. As a robustness check, I use Line fixed effects instead of Line × Section fixed effects in Table A.4 (Appendix) and show that worker characteristics are balanced across HD and LD sections as well.

Table 4: Randomization Check

	Panel A: Outcomes relevant at work						Panel B: Other outcomes		
	Tenure	Muslim co-workers Hindus	Orders	Communication	Age	Schooling	Trust	Altruism	Inter-religious contact outside work
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mixed × LD	0.0886	0.0269	0.0482	0.0894	1.532	-0.219	0.537	0.0203	0.0366
	(0.342)	(0.0218)	(0.0556)	(0.0540)	(1.400)	(0.497)	(0.354)	(0.224)	(0.0466)
Mixed × HD	0.0142	0.0169	0.000148	-0.0350	0.804	0.330	-0.0732	-0.0563	0.00739
	(0.329)	(0.0172)	(0.0470)	(0.0484)	(0.822)	(0.364)	(0.302)	(0.166)	(0.0477)
p(Mixed × LD = Mixed × HD)	0.860	0.721	0.477	0.076	0.621	0.338	0.156	0.755	0.635
Mean Dep Var.	4.45	0.12	0.73	0.53	34.47	7.84	3.79	6.65	0.452
Line × Section Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Religion Effects	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	586	478	586	586	586	586	586	586	586
Adj. R^2	0.121	0.028	0.011	0.043	0.074	0.600	-0.007	-0.014	0.097

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the section-team level. "Tenure" and "Schooling" are measured in years and grade respectively. "Orders" denotes whether a worker is comfortable taking orders at work from non-coreligionists. "Communication" denotes whether an individual is equally comfortable communicating with coreligionists and non-coreligionists. Survey questions on “Trust" and “Altruism" are used from the World Value Survey (WVS). The dependent variable "Interact-Cross" refers to the cross-religion interaction of workers at baseline, outside of work.

3.4 Checks for dependency sorting

Since the majority of workers continued to work in their original sections (i.e. the area of work was not randomized), one might worry about distinguishing between the effects of task types versus worker types (on team productivity) in religiously diverse teams. This is particularly important if workers are able to self-select into High or Low Dependency sections. The randomization check already rules out such systematic sorting. Nevertheless, I address this concern in more detail in Appendix A.2 where I argue that balance in worker characteristics across HD and LD tasks is due to the hiring policy at the firm and not simply by chance. The HR manager always has a pool of job applicants who are called upon on a first-come-first-serve basis, as and when vacancies become available. As a result, workers do not have the option to choose their area of work when they join. However, workers may quit at a different rate across the two types of sections, leading to possible selection bias. If that were the case, this would be reflected in the average tenure of workers in HD and LD sections. As shown in Table A.4 (Appendix), this is not the case - tenure is balanced between workers in HD and LD sections. I then show that only a handful of workers have switched their area of work from when they first joined the firm. Finally, I show that these switches are not correlated with observable characteristics of the workers and have happened purely due to organizational requirements of the firm.

4 Econometric specification

Outcomes in this paper are measured at three levels: 1. Production line-level 2. Production line-section-level and 3. Individual-level. Line-level real output data is linked to the firm's revenues. Line-section-level ratings were recorded by production supervisors daily during the period of the experiment only. These data help investigate the source of line-level differences in real output. Survey measures at baseline and endline are at the individual worker level. I use these to study actual worker interactions and treatment effect on preferences.

Line-Level specification

I compare line-level output between HD-Mixed lines (LD Non-mixed) and LD-Mixed lines (HD Non-Mixed) as shown in Figure 3. The specification used is:

$$Y_{klst} = \beta_1 T_k + \alpha_l + \alpha_s + \alpha_t + \epsilon_{klst} \tag{1}$$

where Y_{klst} is output from line-cohort-level team k, in line l, in shift s on day t and T_k denotes the treatment status (1 if HD-Mixed and 0 if LD-Mixed). The coefficient β_1 denotes the line-level treatment effect. α_l, α_s and α_t are line, shift and day fixed effects respectively. I include production line fixed effects to control for product type, shift fixed effects to account for differences in worker productivity at different times of the day (morning, afternoon, night) and day fixed effects to control for factory-wide shocks. Standard errors are clustered the production line-cohort-level. Since there are only 15 clusters at the line-level, I also present wild cluster bootstrap standard errors (Cameron et al., 2008) for these regressions.

Line-Section-Level specification

Each day, supervisors assigned a rating (out of 5) to each section of each line based on team performance on that day. I use this data to evaluate the source of line-level differences in output. The following baseline specification is used:

$$Y_{mklst} = \beta_1 Mixed_{mkl} \times LD_{ml} + \beta_2 Mixed_{mkl} \times HD_{ml} + X_{mkl} + \alpha_{ml} + \alpha_s + \alpha_t + \epsilon_{mklst} \tag{2}$$

where Y_{mklst} is the performance rating of section m of team k in line l in shift s on day t. $Mixed_{mkl}$ denotes if that section has a religiously mixed or homogeneous team (which is determined by the line level team type k). LD_{ml} and HD_{ml} are dummies coded 1 if the section is classified as a HD or LD respectively (this is defined by line l and section m only). I use the interaction terms $Mixed_{mkl} \times HD_{ml}$ and $Mixed_{mkl} \times LD_{ml}$ to identify effects of having mixed

teams in HD and LD sections respectively. Since *line* × *section* effects α_{ml} are included in these regressions, the dummies HD_{ml} and LD_{ml} are not separately introduced. X_{mkl} denotes a vector of section level controls such as the average schooling and tenure of workers. α_s and α_t are shift and day fixed effects respectively.

Individual-level specification

I surveyed workers both at baseline and endline. I use baseline data for randomization checks as shown in section 3 and also for heterogeneous treatment effects which follow in section 5; and endline data to evaluate treatment effects on preferences of workers. The main specification is:

$$Y_{imkl} = \beta_0 + \beta_1 Mixed_{mkl} \times LD_{ml} + \beta_2 Mixed_{mkl} \times HD_{ml} + X_{imkl} + \alpha_{ml} + \epsilon_{imkls} \tag{3}$$

where Y_{imkl} is the outcome of interest of individual worker i of section m of team k in line l. X_{imkl} are individual level controls. All other variables are described exactly as before. The treatment effects are estimated by coefficients just as in line-section level specifications described above.

5 Results

5.1 Production data

5.1.1 Line-Level

The production supervisors record total output from each production line at the end of each shift. Output is typically measured as the total number of pieces produced in the shift. Table 5 compares line-level output of HD-Mixed lines versus LD-Mixed lines, as described in Figure 3. Observations in this regression are at the line-cohort-day-level. The outcome variable in

Column (1) is the log of total output produced by a line-level team in a shift. Column (1) in Table 5 shows that HD-Mixed lines on average produced 4.9 log points (5.3%) lower output compared to LD-Mixed lines, over the period of the intervention. This effect is economically large. Given average output per shift of 450,000 pieces (across all lines) and the average product priced at Rs 10 ($ 0.13), this suggests that the firm's revenue would increase by up to Rs.225,000 ($3100) per shift, from having lines with only LD mixed sections compared to having lines with only HD mixed sections.

Table 5: Treatment Effect on Line Level Output

	(1) **Log Output (Pieces)**	(2) **Total Output (Boxes)**
HD-Mixed Line (LD Non-Mixed)	-0.0487*** (0.0163)	-100.4*** (32.56)
Bootstrap (Wild Cluster) C.I.	[-0.093, -0.013]	[-190.1, -11.11]
Day Effects	Yes	Yes
Shift Effects	Yes	Yes
Production Line Effects	Yes	Yes
Mean Dep Var.	10.80 (1.24)	1564 (1450)
N	1045	1045
Adj. R^2	0.722	0.639

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the production line-cohort-level in parenthesis. HD-Mixed Line (LD Non-Mixed) is a dummy coded 1 for a line-level team with all HD sections religiously mixed and LD sections non-mixed, and 0 for exactly the opposite line-level structure. Wild cluster bootstrap (Cameron et al., 2008) confidence intervals in [] brackets.

The firm also records total output using the number of boxes of products that were packed at the end of the production line in a shift. These boxes are used to ship products to the market and each box typically includes multiple pieces of a product. The effects are robust to using this variable as the outcome instead. Since each production line can manufacture more than one variant of the same product, I show robustness to including *line* × *variety* fixed effects in Table

C.9 (Appendix). Finally in Table C.10 (also in the Appendix), I include *line* × *day* fixed effects – the results remain robust.

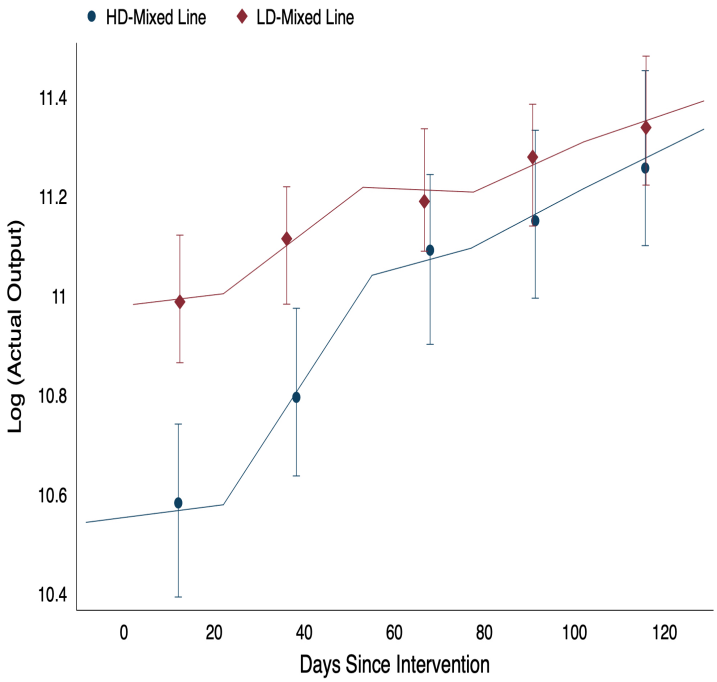
Based on raw material usage, supervisors at the firm record standard (expected) output against actual output produced for each line in each shift. Negative deviations from standard output imply greater raw material wastage. In Appendix B, I show wastage or "Output Gap" is larger in HD-Mixed lines despite raw materials being allocated equally among HD-Mixed and LD-Mixed lines. Furthermore, variance of Output Gap is also greater in HD-Mixed lines, suggesting that mixing workers in HD sections causes greater uncertainty in terms of achieving daily output targets.

Over the entire period of the intervention HD-Mixed lines produced 5.3% lower output than LD-Mixed lines, but how did the treatment effect evolve over time? This would inform us whether repeated interaction with the same set non-coreligionist coworkers can help ameliorate some of the negative effects on output. In Figure 4, I present an event study plot of actual output (log) produced over the period of the intervention, by team type. These are from binned regressions using the same specification as in section 4, with the treatment period split into 5 equal sized bins. The difference in actual output produced by the two types of line-level teams was the largest at the beginning of the intervention and gradually attenuated over time. The difference in standard output between the teams is much smaller (and statistically indistinguishable from 0) as observed in Figure D.8 (Appendix D). This would be expected if the firm did not react to these differences across teams by redistributing output away from low productive lines to high productive ones. Both actual output and standard output followed an upward trajectory throughout the 4 months of the experiment. This could be because of two possible reasons: 1. The firm itself was adjusting to new teams and therefore only gradually increased expected output as workers became more comfortable with each other and 2. The experiment was timed to coincide with the period during which the factory faces high demand for its products; so that production remains uninterrupted, absenteeism is low and teams don't disintegrate. It is

difficult to point out which of these factors was more important.

Overall, these results imply that religious diversity lead to greater output loss in HD sections compared to LD sections. However, they are not informative of whether having mixed teams in LD sections only, would also cause output loss relative to homogeneous teams. This analysis is not possible at the line-level because there is no homogeneous line-level team by design. Therefore, I take this up in the line-section-level analysis, where such comparison is possible due to the presence of teams composed of only Hindu workers. It will help identify the source of line-level differences in output (sections that drive overall differences at the line-level) reported above, and allow us to understand mechanisms behind these findings, by studying heterogeneous effects (based on worker characteristics) using more granular data.

Figure 4: Treatment Effect on Output (Event Study)



Note: This figure is generated from binned regressions[a] using exactly the same controls variables as in Table 5. The treatment period is divided into 5 equal sized bins. The outcome variable is actual output produced (logged).

[a]This plot is created using the STATA command binsreg, which implements binscatter estimation with robust inference proposed in Cattaneo et al. (2019).

5.1.2 Line-Section-Level

I now present treatment effects on line-section level ratings. Recall that there are 4 different types of teams at this level - HD Mixed, HD Non-Mixed, LD Mixed and LD Non-Mixed. The performance of each section was rated (between 0 to 5) daily by production supervisors. These ratings were based on a benchmark measure of time-use efficiency. The benchmarks were of course different across sections. For example, Mixing sections were rated on the number of batches mixed per hour, while most other sections downstream until Packing were rated on number of trays with unfinished products that were sent onto the following section every hour, accounting for the number of trays received from the previous section. This ensured that no section was penalized for the actions of sections upstream. Packing sections were rated on number of boxes completed and on packaging material wastage.

Table 6 presents the baseline results from the line-section-level analysis. In column (1), I regress raw ratings on a dummy variable (Mixed) that denotes whether a line-section-level team is religiously mixed or not. The coefficient on Mixed is negative and marginally significant suggesting that mixed teams on average receive lower ratings. Note that line × section effects are included in all specifications in the line-section-level analysis, whereby the identifying variation comes from within the same line-section across different treatment cohorts (teams). These are important to include due to the different benchmarks used to rate each section.

All regressions also include the average tenure and schooling of workers in the section, to account for differences between Hindus and Muslims on these dimensions.[14] In column (2), I introduce the interaction terms (Mixed × HD) and (Mixed × LD) to estimate the effect of having a mixed team in a HD section separately from a LD section. The coefficient on Mixed × LD in column (2) is small and not statistically significant while that on Mixed × HD is negative and statistically significant at the 5% level. This suggests that having mixed teams lead to lower rat-

[14] All results are robust to the exclusion of these controls as reported in Table C.12 in the Appendix. I continue to include these controls in the line-section-level analysis, but all results are robust to them being excluded.

ings in HD sections but not in LD ones. In columns (3) and (4), the outcome variable is coded 1 if the the rating received is above median and 0 if lower.[15] The effects with a binary dependent variable are similar to those with raw ratings and more precisely estimated. In summary, this is direct evidence that lower output in HD-Mixed lines (relative to LD-Mixed lines) is caused by lower output in religiously mixed HD sections. In Table C.11 (Appendix), I run separate regressions for HD and LD sections and find similar effects.

Table 6: Treatment Effect on Section Ratings

	Rating (Raw)		**Rating > Median**	
	(1)	(2)	(3)	(4)
Mixed	-0.0204*		-0.0254***	
	(0.0119)		(0.00899)	
Mixed × LD		-0.0067		-0.0047
		(0.0144)		(0.0121)
Mixed × HD		-0.0349**		-0.0474***
		(0.0185)		(0.0121)
p(Mixed × HD = Mixed × LD)		0.229		0.011
Mean Dep. Var.	3.82	3.82	0.44	0.44
	(0.83)	(0.83)	(0.50)	(0.50)
Education and Tenure Controls	Yes	Yes	Yes	Yes
Day Effects	Yes	Yes	Yes	Yes
Shift Effects	Yes	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes	Yes
N	6909	6909	6909	6909
Adj. R^2	0.600	0.600	0.358	0.358

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-level team level. "Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified in columns (2) and (4). Education and tenure control for the mean of schooling and tenure of workers in the line-section-level team.

I next examine whether there is convergence in line-section-level performance over time in

[15] A large fraction of ratings is concentrated between 4 and 5 (see Figure D.10 in the Appendix) especially towards the end of the experimental period when productivity differences reduced across teams. Thus a binary variable might also be appropriate in this setting.

Table 7. This is likely given that line-level output differences between treatment arms attenuate over time. I split the experimental into five equal sized bins (exactly as in the line-level analysis), and show that ratings between mixed and non-mixed teams converged in HD sections. The baseline effect is reported in column (1), where a large negative and statistically significant effect of having a mixed team can be observed. In column (2), I introduce interactions with day bins. Coefficients on earlier bins are larger and they gradually reduce in magnitude. This suggests that the largest negative effect was at the beginning of the experiment when the new teams were first formed; and ratings between mixed and non-mixed teams gradually converged over the intervention period. In columns (3) and (4), the baseline effect and the interactions with day bins are presented respectively for LD sections. The baseline effect is small and not statistically significant, while the interactions are noisy with no clear dynamic pattern.

These results are consistent with two important observations that are formalized in theoretical framework (in section 6). First, having mixed teams in LD sections do not lead to lower output. This lends support to the idea that it is indeed the nature of contact (in HD sections) determined by the type of production process that drives these effects, and rules out the possibility that lower output in HD mixed sections is simply due to the fact that Muslims are less productive individuals. If that were true, LD mixed teams should be less productive too.

Second, the model shows that in HD sections, Muslims are likely to exert higher effort than optimal initially, to prove to the majority group (Hindus) that they are equally hardworking individuals. As a result, Hindu workers would update their priors about Muslim co-workers' effort level upwards, inducing them to put greater effort over time. This is likely to cause the output loss from having mixed teams in HD sections to attenuate, leading to overall differences in output at the line-level being bridged. These patterns are indeed observed in both the line-level and line-section-level output data. In section 6.3, I conduct further empirical tests of this particular mechanism. I study how priors of Hindus at baseline affect overall productivity and whether the degree of attenuation over time depends on these beliefs at baseline.

Table 7: Treatment Effect on Section Ratings: Event Study

	Raw Ratings			
	HD Sections		LD Sections	
	(1)	(2)	(3)	(4)
Mixed	-0.0496***		-0.0005	
	(0.0184)		(0.0142)	
Mixed × 0-25 days		-0.105*		0.0525
		(0.0675)		(0.0615)
Mixed × 26-50 days		-0.0716**		-0.103
		(0.0355)		(0.0724)
Mixed × 51-75 days		0.0279		-0.0134
		(0.0340)		(0.0400)
Mixed × 76-100 days		-0.0647**		0.0579*
		(0.0319)		(0.0286)
Mixed × 101-120 days		-0.0247		-0.0542
		(0.0532)		(0.0446)
Mean Dep. Var.	3.85	3.85	3.80	3.80
	(0.68)	(0.68)	(0.64)	(0.64)
Education and Tenure Controls	Yes	Yes	Yes	Yes
Day Effects	Yes	Yes	Yes	Yes
Shift Effects	Yes	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes	Yes
N	3466	3466	3443	3443
Adj. R^2	0.609	0.609	0.595	0.596

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-level team level. "Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified. Education and tenure control for the mean of schooling and tenure of workers in the line-section-level team.

The final result in this sub-section is an important robustness check. Here, I address concerns regarding potential bias that could stem from not being able to randomize workers into HD and LD tasks. In the randomization check, I show that characteristics of workers are balanced across the four different types of teams at the line-section-level. In addition, in section 3, I also present evidence that workers are not able to sort across jobs in the factory over the period of their tenure, and that observable worker characteristics do not predict the very few

switches that have taken place; these have occurred simply due to organizational requirements of the factory. The fact that average tenure of workers is not different between in HD and LD sections shows that workers do not quit at different rates across these jobs, ruling out selection at the time of appointment as well. However, one might still worry that it is in fact not the interaction between religious mixing and HD tasks itself that leads to productivity loss, but some third factor which interacts differently with religiously mixed teams in HD sections relative to LD sections. For example, it might be the case that differences in schooling between Hindus and Muslims is not important in LD sections, but it might be a problem in HD sections, given the nature of contact. In other words, it is differences in education between Hindus and Muslims that matters in some tasks and not others, as opposed to the production technology and the nature of contact being the important factor.

To deal with this, I introduce interactions between the dummy Mixed and these variables as controls, in addition to the interaction terms Mixed × HD and Mixed × LD in the line-section-level analysis. I specifically use three variables: group size, tenure of workers and schooling of workers. HD sections tend to have more workers, and one might be concerned about differences in responses of workers from being mixed in larger groups as opposed to smaller groups. For example, diversity might be costly when groups are larger because there is likely to be a wider set of issues that require coordination on, leading output loss in HD sections but not in LD sections. The other two are more obvious choices given the differences amongst Hindu and Muslim workers on these dimensions.

The results are reported in Table 8 - I introduce the interacted controls sequentially. Column 4 reports the specification with all the controls. Reassuringly, the interaction term Mixed × HD remains negative and significant after the inclusion of these factors. Note that the interactions Mixed × Schooling and Mixed × Tenure are both positive and meaningful in magnitude (though not statistically significant). This suggests that higher tenure and schooling can dampen some of the negative effects of having religiously mixed teams in HD sections. Note also in columns

(3) and (4), the coefficients on the interaction Mixed × LD are statistically significant suggesting diversity might be costly in LD tasks as well, if workers have very low tenure or schooling.

Table 8: Treatment Effect on Section Ratings: Adding key controls

	Raw Ratings			
	(1)	(2)	(3)	(4)
Mixed × LD	-0.00674	-0.0210	-0.122**	-0.173***
	(0.0144)	(0.0179)	(0.0507)	(0.0603)
Mixed × HD	-0.0349**	-0.0609**	-0.164***	-0.216***
	(0.0185)	(0.0264)	(0.0522)	(0.0614)
Mixed × Group Size		0.00396	0.00784**	0.00670*
		(0.00335)	(0.00377)	(0.00400)
Mixed × Tenure			0.0169**	0.0119
			(0.00748)	(0.00854)
Mixed × Schooling				0.0102
				(0.00717)
p(Mixed × HD = Mixed × LD)	0.229	0.112	0.069	0.075
Mean Dep. Var.	3.82	3.82	3.82	3.82
	(0.83)	(0.83)	(0.83)	(0.83)
Education and Tenure Controls	Yes	Yes	Yes	Yes
Day Effects	Yes	Yes	Yes	Yes
Shift Effects	Yes	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes	Yes
N	6909	6909	6909	6909
Adj. R^2	0.600	0.600	0.600	0.600

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-level team level. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified. Education and tenure control for the mean of schooling and tenure of workers in the line-section-level team.

5.2 Endline phone survey

The endline survey focused on two main sets of outcomes: 1. Those that capture actual interactions between workers during production and 2. Preference towards non-coreligionists at work (communication and co-working). Only a phone survey could be conducted because of restrictions related to COVID-19. As a result, a large set of outcomes that I was interested in, including

political preferences, could not be recorded. I take up each of the two sets of survey outcomes in turn.

Mechanisms: Worker interactions

In Table 9, I focus on the first set of factors. These collectively proxy for the degree of cohesion and coordination in a line-section-level team. There are three main outcomes variables. The first question asked workers to point out free riders in their team. If a worker identifies his teammate to have not contributed to the team as much as other workers did, or to the extent that is expected, then the worker is asked to identify that particular teammate as a Free Rider. I then asked workers to identify teammates who have blamed them in the past for not performing up to the mark. The outcome variable Blamed is a dummy variable coded 1 for teammates who have blamed the respondent at least once, during the intervention period. The final question asked workers to pick teammates who they would give up their Relief Time for, if asked or already have in the past. Relief Time refers to breaks that each worker is entitled to at regular intervals during their shift. Typically in HD sections workers need to coordinate on breaks to a greater degree than in LD sections. Note that these questions were asked retrospectively in lieu of more high frequency data, since many workers reported to have had problems in the past but also mentioned that they subsided over time.[16]

Observations in Table 9 are at the worker-teammate level for line-section-level teams. In other words, there are $(N-1)$ observations for each worker, where N denotes the total number of workers in the line-section-level team. I include Line × Section fixed effects and therefore compare similar size teams doing the same task. Columns (1), (3) and (5) show that Mixed teams perform worse on all of these measures. Workers in mixed teams are 3.8 percentage points (29%) more likely to identify a teammate as a free rider, 3.7 percentage points (46%) more likely to have been blamed by a teammate and 6.2 percentage points (21%) less likely to

[16] For example, workers were asked if they have been blamed by a teammate at least once in the past, or asked to identify workers who would free-ride at the beginning of the intervention.

give up their relief time for a teammate. In columns (2), (4) and (6), I introduce the interaction terms Mixed × HD and Mixed × LD to test for differential effects by task type. It can be observed that having mixed teams lead to greater frictions in HD sections.

Table 9: Treatment effect on worker interactions

	Identified teammate as Free Rider		**Blamed by teammate**		**Unwilling to give up Relief Time**	
	(1)	(2)	(3)	(4)	(5)	(6)
Mixed	0.0383***		0.0368**		0.0623*	
	(0.0126)		(0.0145)		(0.0332)	
Mixed × LD		0.0286		0.0737***		0.0315
		(0.0201)		(0.0204)		(0.0439)
Mixed × HD		0.0408***		0.0276*		0.0710**
		(0.0144)		(0.0162)		(0.0394)
p(Mixed × HD = Mixed × LD)		0.604		0.053		0.478
Mean Dep. Var.	0.13	0.13	0.08	0.08	0.30	0.30
Worker skill F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Religion F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	3964	3964	3934	3934	4028	4028
Adj. R^2	0.014	0.014	0.011	0.012	0.044	0.045

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-team level. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified in columns (2), (4) and (6). Observations are at the section worker-teammate level i.e. there are (N-1) observations per worker, where N denotes the number of workers in the section. Workers were asked to choose teammates who they: (1) think have been a free-rider since teams changed (2) they have been blamed by during the experiment and (3) would give up their relief time for.

Surprisingly however, I find that workers in LD-Mixed sections report to have been blamed more by co-workers than those in HD-Mixed sections. Note that both of these effects are statistically significant on their own. It is likely that under-performance in LD sections is attributable to individuals to a greater degree, which is perhaps why we observe this pattern. [17] More gen-

[17] It is also plausible that by endline these frictions had subsided more in HD sections than in LD ones.

erally, it can be observed that mixed teams in LD sections also suffer from these frictions to a greater extent than homogeneous teams - the effects on the interactions Mixed × LD are positive and meaningful in magnitude though not precisely estimated. In fact, one cannot statistically reject that the effects in HD and LD sections are different, though the effects in HD sections are much larger. Importantly however, these do not translate into mixed teams in LD sections performing any worse than non-mixed teams, unlike in HD sections, as shown in the line-section-level output results. The sample is restricted to only Hindu respondents in Table C.15 (Appendix C) and similar patterns are observed.

Overall, these results are consistent with the treatment effects on output and help elucidate mechanisms behind those findings. They point to the fact that ethnic identify is salient in HD sections which have a tight-knit production structure where worker efforts are strong complements. Priors about out-group members are likely to be more important in these jobs and a low initial prior about their ability could cause frictions of the forms described above. This in turn leads to lower in-group cohesion and coordination leading to output loss. However, despite these frictions, output differences between HD-Mixed and HD Non-Mixed sections attenuated over time. The next set of results study treatment effects on preferences of workers towards non-coreligionists at endline, and formally tests whether repeated contact in a HD setting indeed lead to improved inter-group relations.

Preferences at Endline

I now turn to the results on preferences of workers towards non-coreligionists. Note that in this case, treatment effects are restricted to Hindu workers only, because Muslim workers are always in mixed teams. I use three main outcome variables, two of which are questions also asked at baseline. Workers were asked if they are *equally* comfortable taking orders from non-coreligionists (*Taking Orders*), whether they in general find communicating with non-coreligionists as comfortable as co-religionists (*Communicating*) and finally whether they prefer to be in all Hindu groups if teams were to change again in the future (*Co-working*). While the first two

questions are unincentivized, for the third question, surveyors mentioned to the workers that their responses would be recorded for future team changes, and would also be forwarded to supervisors.

All outcomes show positive effects from mixing (Table 10). I first report the main effect of being randomized into a mixed team. Relative to those in homogeneous teams, Hindu workers in mixed teams are 12.5 percentage points (16.9%) more likely to report that they are comfortable taking orders from Muslims (Column 1) and 9 percentage points (18.36%) more likely to be comfortable communicating with Muslims (Column 2). Finally, Column 3 shows that they are 11.5 (29.5%) percentage points more likely to not express preference for being in a Hindu-only team. These effects are economically significant in magnitude and suggest large gains from repeated contact with Muslim coworkers. In Columns (2), (4) and (6), I introduce the interaction terms Mixed × HD and Mixed × LD. These effects are entirely driven by contact in HD sections. The coefficients on Mixed × HD are economically large in magnitude and statistically significant at the 1% level.

The coefficients on Mixed × LD are small and not statistically significant, suggesting a null effect in LD sections. The difference between the effects in HD and LD sections are large and statistically significant. These findings on positive attitudes of Hindu workers in HD-Mixed sections towards Muslims are consistent with attenuating output differences between HD-Mixed and HD Non-Mixed teams, as well as overall convergence in line-level output of the two different line-level team types.

In Table C.20 (Appendix), treatment effects are broken down by worker skill to show that the largest effects occurred on unskilled workers. This is primarily because they have a lower mean value of outcomes to begin with as shown in Table C.21 (Appendix). To get a sense of the magnitude of the effects, a comparison of pre-treatment means between skilled and unskilled workers suggests that the treatment effect of being in a mixed HD team for unskilled workers was large enough to bring their outcome averages up to those of skilled workers.

Table 10: Treatment Effect on Preferences: Hindus

Comfortable:	**Attitudes towards Muslims**					
	Taking Orders		**Communicating**		**Co-working**	
	(1)	(2)	(3)	(4)	(5)	(6)
Mixed	0.125***		0.0961**		0.115***	
	(0.0448)		(0.0406)		(0.0348)	
Mixed × LD		0.0180		-0.0843		0.0198
		(0.0778)		(0.0615)		(0.0626)
Mixed × HD		0.187***		0.200***		0.169***
		(0.0555)		(0.0407)		(0.0439)
p(Mixed × HD = Mixed × LD)		0.086		0.000		0.0807
Mean Dep. Var.	0.74	0.74	0.49	0.49	0.39	0.39
Sample	Baseline Mean		Baseline Mean		Endline Non-mixed Mean	
Worker skill F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	448	448	448	448	448	448
Adj. R^2	0.071	0.078	0.064	0.083	0.055	0.058

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-team level. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified in columns (2), (4) and (6). “Orders" refers to whether a worker is comfortable listening to orders from a non-coreligionist. “Communication" refers to whether a worker is comfortable communicating with non-coreligionists in general. Co-working refers to whether a worker is comfortable being put in a team with a non-coreligionist in future.

Overall, it is insightful and non-obvious that the largest positive effects of treatment on preferences occurred in teams that also suffered the largest negative output shocks. This suggests that working in close quarters even with some frictions (HD teams) leads to more positive effects on group relations than working in LD teams. These results emphasize the importance of contact that forces people to learn to work together in overcoming existing differences leading to reduced intergroup prejudice. Purely from a profit maximizing point of view however, firms will have no incentive to mix workers in HD tasks if it leads to output loss. This unfortunately

suggests that discrimination is likely to persist in equilibrium unless management practices are specifically in place to mitigate them.

6 Model

This section presents the theoretical framework. The primary objective of the model is to rationalize the core empirical results, especially the attenuation in output losses in HD-mixed sections over time. The model makes predictions specially with respect to the mechanism that leads to this attenuation, as well as heterogeneous treatments effects based on worker characteristics, which I subsequently test for in the data.

A key distinction is made between Hindu and Muslim workers in this framework. Consistent with majority-minority relations, a large section of Hindu workers at the factory have never worked with Muslims in the past, while 100% of the Muslim workers have. Based on this asymmetry in exposure at baseline, together with the evidence on discrimination against Muslims in access to education and labor markets in India (Kalpagam et al., 2010; Basant, 2007),[18] I assume Hindus (mistakenly) believe that Muslims are not as hardworking as them.[19] Muslim workers do not make this distinction between in-group and out-group workers given that they have had much greater contact with the majority group. This majority-minority asymmetry leads to multiple equilibria in HD interactions due to complementarities in the production function, while in LD it does not matter.

The lack of past exposure with Muslims means that Hindus do not fully understand the game, they are not forward looking and act simply based on their existing priors. As a characteristic feature (not strategic) however, their priors are updated by exposure to Muslims workers. Muslim workers on the other hand are forward looking given their past experiences of work-

[18] In fact, Muslims in my sample have significantly lower schooling than Hindus (Table C.8).

[19] An implicit assumption here of course is that in reality Hindus and Muslims are equally productive. I show direct evidence of this for workers in section 7 and for operators and supervisors in Appendix E.

ing with Hindus; they are aware of the bias Hindu workers have and also understand that it is possible for them to updated.

6.1 Setup

6.1.1 Production functions

Suppose production is composed of two types of tasks, HD (High Dependency) and LD (Low Dependency). Total output in a production line is a function f of output y_{HD} in HD tasks and y_{LD} in LD tasks.

$$f(y_{HD}, y_{LD}) \tag{4}$$

Suppose there are a total of N_{HD} workers in a HD section and N_{LD} workers in a LD section. The cost of high effort is greater than low effort for each worker: $c(e_H) > c(e_L)$. Output is stochastic[20] - the effort levels of workers determine the probability of observing High versus Low output. In HD sections, the joint effort of all workers determine the probability of high (O_H) or low (O_L) team output. Therefore expected output in a HD section is given by:

$$\mathbb{E}(y_{HD}(e_i, ..., eN_{HD})) = p(e_i, ..., eN_{HD})O_H + (1 - p(e_i, ..., eN_{HD}))O_L \tag{5}$$

where E_{HD} is a vector of individual efforts (e_1,....,$e_{I_{HD}}$) with $p_e > 0$. O_H and O_L denote high and low output levels respectively. In LD sections, total output is the sum of individual expected output:

$$\mathbb{E}(y_{LD}(e_i, ..., e_{N_{LD}})) = \sum_{i=1}^{N_{LD}} \{p_{e_i} o_h + (1 - p_{e_i}) o_l\} \tag{6}$$

[20] There are several factors that workers do not directly have control over, which could influence productivity on a particular day. These include, but are not restricted to: inadequate raw material planning by supervisors, machine breakdowns that cannot be foreseen and inefficient handover of production from supervisors in the previous shift.

6.1.2 Binary effort environment

For simplicity (without loss of generality), consider a binary environment with two effort choices for workers: e_H (High) and e_L (Low) respectively. Assume also that there are only 2 workers in each production section. The probability of observing high output O^H given the effort choice of each worker is as follows:

$$(e_H, e_H) = p_H \tag{7}$$

$$(e_H, e_L) = (e_L, e_H) = p_{HL} \tag{8}$$

$$(e_L, e_L) = p_L \tag{9}$$

Definitions and Assumptions

1. $p_H > p_{HL} > p_L$ with $p_H > \frac{1}{2}$ and $p_L < \frac{1}{2}$

2. $c(e_H) > c(e_L) = 0$ (cost of low effort is normalized to 0)[21]

3. In HD sections, $(p_H - p_{HL})(O^H - O^L) > c(e_H) - (p_{HL} - p_L)(O^H - O^L) > 0$. [22]
   Workers prefer a high steady state (expected) output to a low steady state (expected) output since it affects their probability of promotion. However, conditional on low effort from their teammate, exerting low effort is optimal in each period.

4. In LD sections the analogue to the point above implies that expected utility from high effort is greater than low effort, i.e. $(p_H - p_{HL})(O^H - O^L) > c(e_H)$.[23] In LD, high effort is therefore the dominant action.

5. Workers do not directly observe effort levels of their teammates, they simply observe realized output in each period.

[21] Note that the probabilities in 1 and effort costs in 2 do not vary by religion. This implies Hindus and Muslims are equally productive.

[22] This expression is obtained by re-writing: $p_H O^H + (1 - p_H)O^L - c(e_H) > p_L O^H + (1 - p_L)O^L > p_{HL} O^H + (1 - p_{HL})O^L - c(e_H)$.

[23] The expression is $p_h o^h + (1 - p_h)o^l - c(e_H) > p_l o^h + (1 - p_l)o^l$.

6. There are no worker “types" within groups. All Hindus are the same and all Muslims are the same.

6.2 Analysis of the model

Workers interact repeatedly for T periods. HD sections are the interesting case here due to complementarity in worker efforts in the production function. I first solve the model for HD sections and subsequently discuss LD sections.

6.2.1 Hindu workers (majority group)

Hindu workers who are in mixed teams could have been in non-mixed ones, which they believe would be more productive. In other words they assign probability p_t (in period t, initial belief is p_0) (exogenously given) on their Muslim teammate exerting high effort, which in the case of coreligionists (other Hindus) is 1.

A Hindu worker's problem is given by:

$$V = max_{(e_t)_{t=0}^{T}} \sum_{t=0}^{T} P_t(e_t, p_t) \tag{10}$$

where e_t denotes the action in period t (choice variable), p_t is the prior at t (state variable) and $P_t(e_t, p_t)$ is the expected (perceived) payoff.

Each period, given their current prior, their own action and realized output, the Hindu worker's belief about the effort level of their Muslim teammate is updated. The transition matrix at any period t, with current prior p_t, is shown below. Note that Hindus are not strategic about this, belief updating is simply a characteristic they have, which is triggered upon contact with Muslims. They simply act on their current priors.

Table 11: Bayesian Updating (Hindu workers) - Prob(Muslim worker exerts high effort)

Own Effort/Realized Output	O^H	O^L
e_H	$\frac{p_t p_H}{p_t p_H + (1-p_t)p_{HL}} > p_t$	$\frac{p_t(1-p_H)}{p_t(1-p_H)+(1-p_t)(1-p_{HL})} < p_t$
e_L	$\frac{p_t p_{HL}}{p_t p_{HL} + (1-p_t)p_L} > p_t$	$\frac{p_t(1-p_{HL})}{p_t(1-p_{HL})+(1-p_t)(1-p_L)} < p_t$

Note: The prior of a Hindu worker in period t is denoted by p_t.

The problem for a Hindu worker in any given period t and prior p_t can simply be re-written as a choice between High or Low effort. As mentioned earlier, I assume that net expected payoff in a high effort (from both players) environment is greater than the net expected payoff in a low effort (from both players) environment.

Therefore, expected payoff to Hindu worker h, from putting effort e_H is

$$P^h(e_H, p_t) = p_t\{p_H O^h + (1-p_H)O^L\} + (1-p_t)\{p_{HL}O^h + (1-p_{HL})O^L\} - c(e_H) \tag{11}$$

whereas the payoff from putting effort e_L is

$$P^h(e_L, p_t) = p_t\{p_{HL}O^H + (1-p_{HL})O^L\} + (1-p_t)\{p_L O^H + (1-p_L)O^L\} \tag{12}$$

Comparing (12) and (11) we then see that the Hindu worker's problem is a choice every period based on priors, they put high effort if (11) > (12), which is the case when

$$p_t > \frac{c(e_H) - (p_{HL} - p_L)(O^H - O^L)}{(p_H + p_L)(O^H - O^L)} = \bar{p} \tag{13}$$

From assumption (3) it can be seen that the numerator in the RHS is positive.

6.2.2 Muslim workers (minority group)

Unlike Hindu workers, Muslim workers have always been in mixed teams. They are used to this type of discrimination - they are aware that Hindu workers who have had little contact in the past with Muslims, believe with probability $p < 1$ that Muslims are not as hardworking.

Muslim workers choose an optimal effort investment path based on the time horizon. At any given time t and set of history s (which determines prior p_t of Hindu teammates), Muslim workers choose an effort level. Their problem can be written as:

$$V = E^{\pi}(\sum_{t=0}^{T} P_t | p_o = p) \tag{14}$$

where π denotes the mapping from a set of histories (from 0 to $t-1$) to actions $a_t = (e_H, e_L)$ and P_t denotes expected payoff in each period conditional on the set of history and preferred action choice in that period. State s_t (history of high vs low output events) defines the current belief (p_t) of the Hindu worker regarding their Muslim co-worker.

Note that for any t=k, the problem above can be re-written as

$$V_k^{\pi}(s) = \{\sum_{s' \in S_{k+1}} P_k(s,a) + p_k(s'|s,a) V_{k+1}^{\pi}(s')\}, k = N-1, ..., 0 \tag{15}$$

where π denotes mapping from each possible history $h_t = (s_o, a_o, ..., s_{t-1}, a_{t-1})$ to actions $a_t = \pi_t(h_t)$. The optimal effort path for a Muslim worker is then a mapping from state histories to actions π^* such that,

$$\pi_k^*(s) \in argmax_{a \in e_H, e_L}\{\sum_{s' \in S_{k+1}} P_k(s,a) + p_k(s'|s,a) V_{k+1}^{\pi *}(s')\} \tag{16}$$

The per-period expected payoff to Muslim workers for efforts e_H and e_L when $p > \bar{p}$ and $p < \bar{p}$ respectively is as follows:

Case 1: $p < \bar{p} \implies$ Hindu worker exerts Low effort

$$e_L : p_L O^H + (1 - p_L) O^L \tag{17}$$

$$e_H : p_{HL} O^H + (1 - p_{HL}) O^L - c(e_H) \tag{18}$$

Case 2: $p > \bar{p} \implies$ Hindu worker exerts High effort

$$e_L : p_{HL} O^H + (1 - p_L) O^L \tag{19}$$

$$e_H : p_H O^H + (1 - p_H) O^L - c(e_H) \tag{20}$$

When $p > \bar{p}$, exerting High effort until T (final period) is the dominant strategy. However, the optimal effort investment path will depend on the horizon of the game when $p < \bar{p}$.

6.2.3 Implications of the model

1. *In a one shot game, if $p < \bar{p}$, the only equilibrium is where both workers exert low effort in a HD section.*

   This trivially follows from assumption (3). For $p < \bar{p}$ Hindus exert low effort. Since (p_H - p_{HL})(O^H-O^L) > $c(e_H)$ - (p_{HL} - p_L)(O^H-O^L) > 0, it is optimal for Muslims to do the same.

2. *Given a long enough horizon (large T), the optimal effort investment path for a Muslim worker would be to exert e_H from the first period. Hindu workers will eventually update $p > \bar{p}$ leading to the transition to higher expected steady state output. The initial negative payoff of the Muslim worker would be compensated for by this transition.*

   A formal proof is provided in the Appendix. The intuition is the following. If a Muslim worker repeatedly exerts high effort (e_H) from $t = 0$, a Hindu worker will find in expectation that there is a larger number of states with high output than would be the case under low effort (e_L) from the Muslim worker. The Hindu worker will thus gradually update their prior about their Muslim teammate's effort level. With T large enough, the Hindu worker's

prior would eventually be high enough (above $\bar{p}$) such that e_H would be the optimal action thereafter. The initial negative payoff to the Muslim worker would be compensated for by the higher expected steady state payoff.

3. *Hindus with high p initially will exert higher effort overall when put in a mixed team relative to Hindus with low p.*

   This naturally follows from the fact that closer a Hindu worker's initial prior is to $\bar{p}$, fewer initial periods with High effort are required from a Muslim worker, before the transition to high expected steady state output is made.

4. *If Hindu and Muslim workers are not differentially productive, there will be no difference in output in mixed versus homogeneous teams in LD sections.*

   This follows directly from the linear production function in LD sections.

This model is a natural way of rationalizing the empirical results in the paper, especially the attenuation in treatment effect on output over time. However, an alternative story could be that Hindu workers update their taste-based preferences about Muslims as opposed to their priors about Muslim workers' effort levels at work (or both). I test for this possibility in section 6.3 and do find some support for it. It is also possible that it takes Hindus longer to learn to coordinate and communicate with Muslims,[24] given their lack of contact at baseline. Nevertheless, the interpretation that the minority group bears the integration cost is borne out by the data. The primary objective of this model is not to rule out other models (specifically taste-based ones) but rather provide one possible framework to rationalize the main findings. The broader idea that Hindus play a series of one-shot games and update their priors gradually while Muslims play a forward looking game due to their past experience of working with Hindus, is useful nevertheless because the additional implications it has are largely consistent with the data, as I show in the next sub-section.

[24] It is unlikely that the ability to communicate (language, dialect etc.) itself acts as a barrier in this context. However, the ease and comfort with each Hindus interact with Muslims might improve over time.

6.3 Testing empirical predictions of the model

The empirical results presented thus far have little to say about how Muslims behave (in HD and LD sections) which is important in fully understanding the nature of the frictions that arise and the role played by each group of worker in mitigating these frictions. In Table C.16 (Appendix), I restrict attention to only mixed teams (across both HD and LD sections) and use dummies for the religion of the respondent, that of the person being referred to in the survey question, as well as their interaction as the main outcome variables. Columns (1), (3) and (5) show that while Muslim workers are more likely to be identified as free riders, be blamed and have less co-workers willing to give up relief time for them, they themselves are less likely to criticize their co-workers. The coefficients on the interactions terms introduced in Columns (2), (4) and (6) show that the criticism of Muslim workers come largely from their Hindu counterparts, while Muslim workers are actually willing to give up relief time for Hindu co-workers with a higher probability (than for Muslim co-workers). This decomposition lends support to the idea that Hindus discriminate against Muslims leading to lower team cohesion and output while Muslims bear the cost of integration in this context, as suggested by the model. In Table C.17, I further decompose the findings of Table C.16 into HD and LD sections and show that the effects discussed above are driven largely by HD sections and less so by LD sections.

Another important implication of the model is that Hindu workers with a high initial p (higher prior that Muslims also exert high effort) are less likely to discriminate against Muslim workers. If this is true, Hindu workers who in the past have had Muslim co-workers and therefore understand that they are equally hardworking, should continue to optimally exert high effort based on their priors, when randomized into a mixed HD team. Thus, mixed teams where Hindus have a higher prior at baseline should perform relatively better than mixed teams in which Hindus have a relatively lower prior at baseline. The line-section-level data can be used to test this by comparing mixed teams in which Hindus on average have had greater contact with Muslims relative to those in which Hindus have had little contact. At baseline, I collected

data on a range of different factors that can help directly test this hypothesis. Most importantly, I have details of the team each worker was in before the intervention, which allows me to determine the share of Muslim workers in their previous team. Tenure at the factory is also a good proxy for past contact with non-coreligionists that workers might have had. In addition to these, I collected data on political preferences of workers at baseline. These specifically relate to factors that could capture anti-Muslim sentiments, such as preference for the Hindu nationalist BJP party and support for the National Registrar of Citizens (NRC) which has been criticized for discriminating against Muslims.

I test how these factors affect the performance of mixed teams in HD sections[25] in Table C.18 (Appendix C). In Panel A, I introduce interactions between the dummy Mixed and average tenure of Hindu workers in the team (Column 1), as well as the share of Muslims co-workers that these Hindu workers had in their team before the intervention (Column 2). The coefficient on Mixed alone is now larger than the average effect presented in Column (1) of Table 7, while the interaction terms are positive. This suggests the negative effect of having a mixed team in HD sections is amplified when Hindu workers in the team have had little past contact with Muslims, which is consistent with the theoretical predictions. The coefficient on the interaction term (Mixed × Share of Muslims in previous team Hindus) though large is imprecisely estimated. This is possibly due to the lack of variation in this measure at baseline. Nevertheless, the magnitude suggests that if we consider a mixed team with Hindu workers who had 40% of their co-workers as Muslims in the past (which is the case in the experiment for mixed teams), the negative effect on performance ratings in a HD section would be small and not statistically significant (β = -0.0232, p-value = 0.709).

In Panel B, I use measures of political preference to capture taste discrimination. Despite the fact that the theoretical framework only considers updating on beliefs about the effort exerted

[25] I also study how these factors affect performance of mixed LD sections. The effects are in the same direction as in HD, but smaller in magnitude and not statistically significant. These results are available upon request.

by Muslims, taste-based preferences could also be an important dimension on which updating occurs. While this is not exactly captured in the model, it is still consistent with the broad framework where Hindus initially exert low effort and eventually update their priors as they learn more about Muslims. The interaction between the dummy variable Mixed and the share of BJP voters as well as the share of NRC supporters are negative, large and statistically significant. This suggests that a larger share of Hindus in a mixed team with these preferences leads to lower performance of these sections. Note that the coefficients on Mixed alone are now difficult to interpret since the share of Hindu workers with preference for the BJP and/or support for the NRC is typically always positive. Importantly, the coefficients on the share of BJP/NRC supporters alone are not statistically significant suggesting that these workers are not more or less productive in general, but under-perform when they have to work alongside Muslims. Finally in Table C.19, I study heterogeneity in the attenuation of ratings in HD sections by inter-religious contact and taste preferences at baseline. I show that mixed teams in which Hindus have either not worked with Muslims in the past or strongly support the BJP/NRC perform relatively worse and the effects do not completely dissipate, while the effect is smaller initially and dissipate entirely if Hindu workers do not have these preferences. These results together show that past contact of Hindu workers with Muslim co-workers help in overcoming some of the frictions that cause productivity losses in HD-mixed sections.[26]

7 Alternative explanations

In this section, I address and rule out alternative explanations for the main results in the paper.

[26] In general, in the baseline data I find that Hindus who have had greater contact with Muslims are less likely to support the NRC or report to favour the BJP. These correlations are consistent with the results in Table C.18.

7.1 Religion and productivity

As reported in Table C.8 (Appendix), Hindus are more educated (higher schooling) and have longer tenure at the factory than Muslims. One might worry that lower output in mixed teams is simply due to lower levels of schooling and tenure amongst Muslims. However, this is unlikely. First, if Muslims were lower productive individuals, one would observe lower output in LD mixed sections too. This is not the case. In Appendix E, I show that even at the supervisor-level, Muslims are not less productive than Hindus on average. Second, in Table C.13 (Appendix), I show that these factors on their own do not determine line-section-level productivity. Nevertheless, I introduce controls for these factors in the line-section-level analysis. Furthermore, in Table 8, I introduce interactions between the dummy denoting whether a team is mixed or not (Mixed) and these factors as control variables, to account for the fact that these differences might only matter in HD sections and not in LD – the results remain robust to these controls.

By design, Muslims workers are only in mixed teams in this experiment. This was done mainly for two reasons. First, Muslims comprise of only 18% of all workers in the factory, whereby forming homogeneous Muslim teams would lead to significant loss of power in estimating the effect of religious diversity on productivity. Second, at baseline, there were no homogeneous Muslim teams to begin with; therefore experimentally generating such teams could raise ethical concerns. The concern this raises is that Muslim workers might be unproductive only in HD tasks and not in LD, which would be consistent with the findings in this paper. This however is extremely unlikely. As shown in Table C.18 there is significant heterogeneity in how mixed teams affect output in HD tasks. When Hindus have been in mixed teams with Muslims in the past, I find that the negative effect of diversity to be muted significantly. Taste preferences of Hindus also play an important role in determining the magnitude of the negative effect. If Muslims generally had lower productivity in HD tasks, it is unlikely that the effects would attenuate so significantly when analyzing heterogeneity by characteristics of Hindu teammates. Secondly, the large attenuation of the main effects over time is also unlikely to be observed if

Muslims on average had lower productivity in HD tasks. In Table C.19, I show that the size of the effect initially and the degree of attenuation over time both depend on the beliefs of Hindus at baseline. This again is unlikely if lower productivity of Muslims was driving this effect. Most importantly, the results in Table 10 (with only Hindus) which show improved preferences of Hindus towards Muslims in HD tasks at endline, do not suffer from this collinearity problem. The results are consistent with the idea that an environment that forces people to learn to work together is important to alleviate group-level differences leading to attenuation of production losses. This further suggests that it is indeed the nature of inter-group contact caused by production function differences that explain these results, as opposed to Muslims simply being less productive at HD tasks.

7.2 New versus old teammates

One might be concerned that the finding that religious diversity negatively affects productivity is driven in part by the difficulty of working alongside new workers in mixed versus non-mixed teams – as opposed to frictions that arise when working alongside non-coreligionists. This could be especially problematic if the share of new people is different across HD and LD sections within mixed teams, relative to non-mixed groups.

I formally reject this possibility in the results reported in Table 12. These are individual worker-level regressions where the outcome variable is the proportion of workers in one's current team (randomized team) who were also in their line-section-team before the randomization occurred. The mean of the outcome variable is 0.34, which is expected since workers in each production line were randomized between 3 different cohorts – whereby roughly a third of the workers would be known to each other after new teams were formed. Importantly, as shown in Columns (1) and (2), the proportion of new workers in each line-section-team is balanced across mixed and non-mixed teams and the interactions Mixed × HD and Mixed × LD are small in magnitude and not statistically significant. This suggests that the findings in this paper do

not simply result from the inability of workers to coordinate with new colleagues, since workers on average had the same proportion of new teammates irrespective of treatment status.

Table 12: Proportion of old teammates after randomization

	Proportion of old teammates (1)	(2)
Mixed	-0.0115	
	(0.0163)	
Mixed × LD		-0.0312
		(0.0321)
Mixed × HD		-0.0006
		(0.0193)
p(Mixed × HD = Mixed × LD)		0.442
Mean Dep. Var.	0.34	0.34
Religion F.E.	Yes	Yes
Line × Section Effects	Yes	Yes
N	577	577
Adj. R^2	0.599	0.600

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-team level. The outcome variable in these regressions is the share of non-coreligionists in each individual's line-section-level team before the intervention. "Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified in column (2).

7.3 Treatment status and section changes due to randomization

The randomization process involved moving 8% of the workers from their original sections (tasks) at baseline so that the line-level team structures in Figure 3 could be achieved. While this is a small share of workers, it is nevertheless important to show that treatment status is not correlated with the probability of section-switching. If that were the case one could argue that the treatment effects are potentially contaminated. For example, if HD mixed teams have a greater share of workers who changed their sections, it is possible that it is in fact the time required to adjust to new tasks that explains the results. To rule this out, I regress a dummy

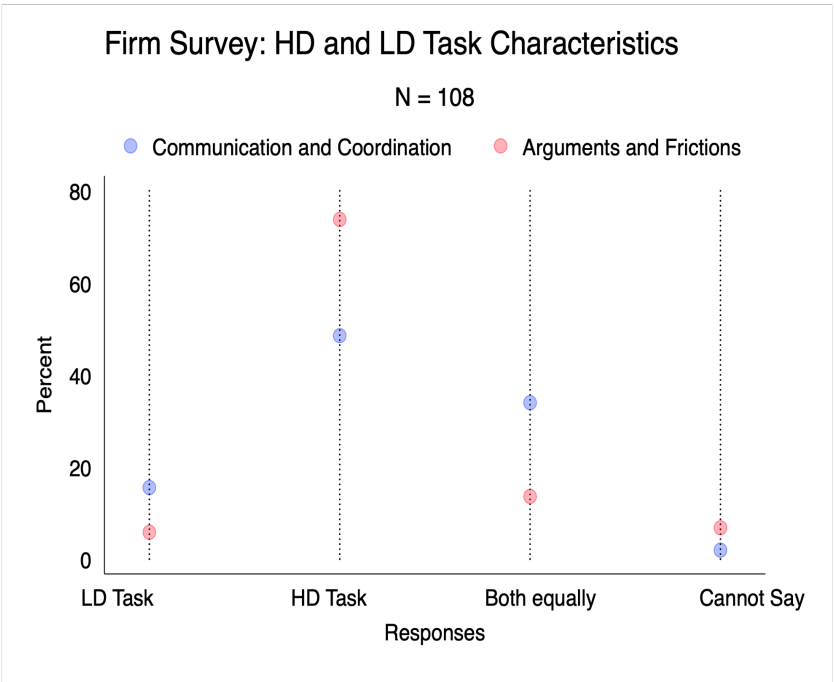
denoting whether the section of a worker was changed due to randomization on the treatment dummies in Table C.22 (Appendix). In columns (1) and (2), I include only a dummy for whether the team is religiously mixed or not (Mixed) and then in columns (3) and (4) I include its interactions with section type (HD or LD). I include line × baseline section effects in columns (1) and (3) and line × section effects in columns (2) and (4). The coefficients across the different specifications are small and not statistically significant. Only in column (4), the coefficient on Mixed × HD is negative and marginally significant, suggesting that the probability a worker switched their baseline section is actually marginally lower in HD Mixed sections. This exercise therefore rules out the possibility that the treatment effects are driven by differential rates of section-switching across treatment arms during the randomization process.

8 Firm supervisor survey

Do firm supervisors understand the costs of diversity and how they may depend on the production function? What management practices do they already have in place or are willing to take in the future to mitigate possible negative effects of diversity? To find out the answers to these questions, I surveyed supervisors and operators (workers that have some leadership role in their job) of 5 different processed food manufacturing firms in April 2021. The survey findings also help reconcile some aspects of the findings from the main experiment. Respondents were shown Figure D.7 with examples of HD and LD tasks. This was followed by a set of questions. Participants were first asked to denote the task (HD or LD) they thought requires: (1) greater continuous coordination and communication amongst co-workers and (2) that is likely to cause more frictions and arguments amongst workers. Figure 5 shows that respondents clearly picked the HD task more frequently for both of these questions. However, while the HD task is chosen by close to 80% of respondents for (2), a fair share of workers picked the LD task or mentioned both HD and LD for (1). This supports the point that LD tasks require a

fair amount of teamwork amongst workers though the degree of instantaneous contact might be less. This pattern suggests that workers are not simply sabotaging or undermining the efforts of out-group members. Rather, one's religious identity assumes more importance in tasks where the propensity of conflict is greater, whereby favoring in-group members likely becomes more important.

Figure 5: Characteristics of HD and LD Tasks

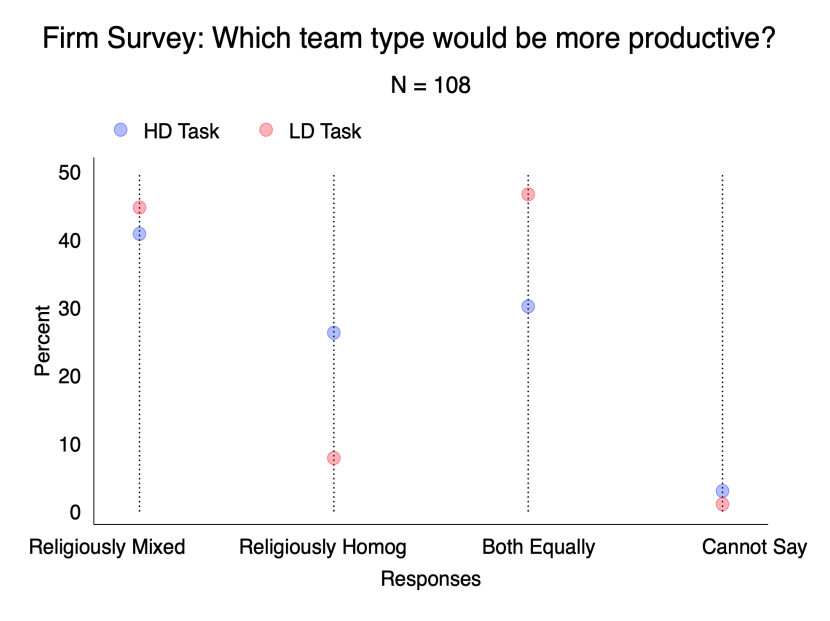


Note: This figure reports the percentage of respondents who picked each option for the following questions. The first question asked respondents to pick the task that they thought requires greater continuous coordination and communication amongst workers (blue dots), while for the second it was the one that is likely to cause more frictions and arguments amongst workers (pink dots).

Next, I asked the participants which type of team (religiously homogeneous or mixed) according to them would be more productive at HD and LD tasks. They were informed that we have already conducted an experiment that estimates these effects, and that they would be rewarded with Rs 25 (about 30% of hourly wage) if their answer matches what we found. The reward money was expected to reduce social desirability bias in the answers. The responses are reported in Figure 6. 40%-45% of workers mentioned religiously mixed teams for both types of

tasks. This could be because of social desirability bias or as I explain next, supervisors actually think of issues beyond direct productivity, that could arise from segregating workers by religion. This could prompt them to answer in this manner. Nevertheless, it can be observed that a significantly higher share of workers mention that a religiously homogeneous team would be more productive in a HD task (30%) relative to a LD task (8%). Overall, this suggests that while 30% of workers did correctly predict how religious diversity would impact productivity in different tasks, the majority did not.

Figure 6: Characteristics of HD and LD Tasks



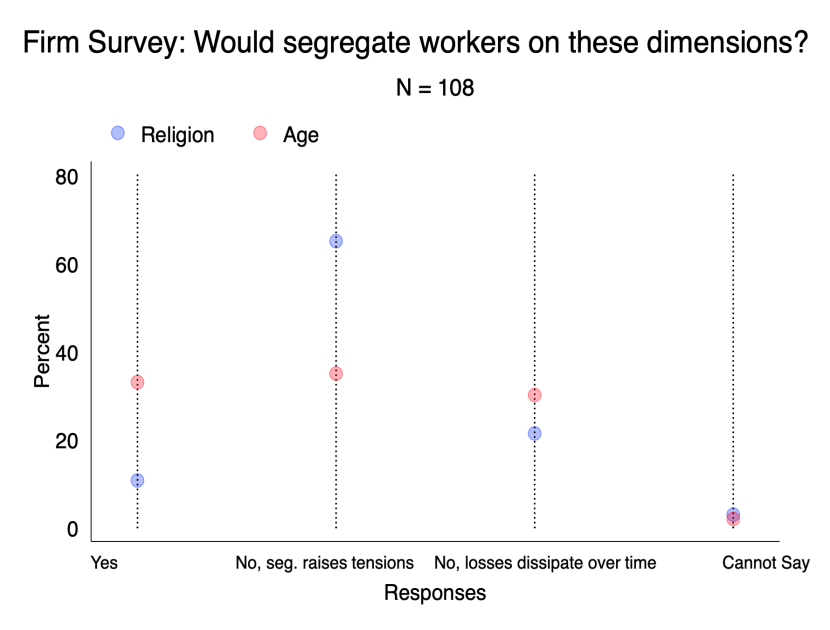
Note: This figure reports supervisors' perception of which type of team (religiously homogeneous or mixed) would be more productive in HD vs LD tasks. It reports percentage of respondents who picked each option when they were asked whether a religiously mixed or a homogeneous team would be more productive separately for HD (blue dots) and LD tasks (pink dots).

While it is possible that most supervisors actually do not understand the costs associated with diversity, and hence do not segregate workers by religion, it is possible that there are costs that do not justify segregation, even if in theory it could improve productivity. To understand whether this is that case more systematically, supervisors were asked if they are willing to segregate workers along religion or age, if workers do not perform well as a team because of their

differences. I use age as a natural benchmark because in the Indian context age differences could also be an important source of conflict amongst teammates.

The results are reported in Figure 7. Majority of supervisors mention that they are unwilling to segregate workers on either dimension, though this share is significantly higher when asked about religion. The most common reason reported for this is that religious segregation is likely to raise tensions further. Informal conversations with supervisors suggest that some of the concerns they have in mind are with respect to tensions that could arise because of such segregation in common areas where workers interact in (canteen, tea room etc.), in addition to tensions on the production floor itself. About a third of the workers also do mention that the negative effects of diversity dissipate with time (as the results from the experiment suggest) which is why they would not segregate workers. The majority though seem to be averse to it due to more direct concerns of conflict.

Figure 7: Characteristics of HD and LD Tasks



Note: This figure presents responses of supervisors when asked if they are willing to segregate workers based on certain demographic dimensions. Percentage of respondents who chose each option for age and religion are denoted by pink dots and blue dots respectively.

The results from this survey provide mixed evidence on whether factory supervisors are actually aware of how religious divisions impact production. However, it is clear that despite the possibility of losses, the majority of supervisors are averse to segregation of workers due to costs that are typically hard to identify as a researcher by simply analyzing production data. This is perhaps why a large number of previous studies find it difficult to reconcile productivity losses arising from diversity with non-segregation of workers in firms they study.

9 Conclusion

This paper shows that the production technology in use at a firm determines the effects of religious diversity on productivity. Greater reliance on non-coreligionist co-workers in High Dependency (HD) environments leads to more inter-group blame and free-riding, lowering team output. In contrast, having diverse teams in Low Dependency (LD) environments do not lead to output loss.

Overall, my findings suggest that both the nature and duration of contact are important in understanding how religious diversity in firms may impact productivity. More importantly, I argue that the conditions under which firms might find it easier to mix workers might not always be those that are also optimal for integration from a social planner's perspective. A tight knit production structure where worker efforts are complements is important to generate an environment where workers find it worth investing in the time and effort to build up social capital that leads to changed preferences towards out-group members - but it might be unprofitable for firms through lost output. This suggests that discrimination could persist in equilibrium since firms that have the ability to integrate workers may not be willing to do so.

Previous work using observational data that rely on frequent team switching to identify the causal effect of diversity on productivity, are either not able to test for the effects of past contact on productivity due to data limitations, or find that past contact does very little to mitigate

the negative effects of diversity (Hjort, 2014). While I cannot directly quantify the cost of team switching within this experiment (I would have to change teams multiple times), my results together with findings from past studies do suggest that in ethnically diverse societies the cost of team switching can be large.

Beyond conceptual contributions, this paper has a few important implications for policy. First, firms which have production function types that resemble HD environments should minimize team switching in order to mediate negative effects of diversity on production. Second, in firms with production function types that resemble LD environments, exposure to non-coreligionists might not necessarily lead to reduced prejudice and discrimination. While this might cost the firm little in terms of lost output, a less cohesive work culture can lead to problems outside of daily production. Such firms might benefit from additional measures to ensure a collaborative environment for workers to interact in. This can even be done outside the workplace for example though sports teams (Lowe, 2020; Mousa, 2018). In general, an open question remains whether such contact outside work can lead to productivity gains at the workplace. Systematic analysis of this possibility can have additional implications for policy – if organized collaborative contact outside the firm does indeed translate into productivity gains within the firm, the cost to productivity in achieving the same inside the firm even if it is for a short period of time, might not be necessary.[27] However, if belief updating with respect to specifics about co-workers' effort levels at work is the driving factor (as suggested in the theoretical framework), contact outside the firm might not be able to entirely mitigate the loss in team production. This could be an important avenue for future research.

As economies undergo structural transformation, the nature of economic production changes, which potentially influences the nature of inter-ethnic interactions. The findings in the paper provide insights on how the cost of diversity might change with this process. My results

[27] Of course arranging such organized collaborative contact might itself be costly for a firm in terms of time and resources.

suggest that it is likely to follow a inverse-U shape pattern as economies move away from traditional agriculture to informal urban activity and eventually to more formal workplaces with regular colleagues. In traditional agricultural societies, land cultivators largely work in LD environments with limited contact with new people, but informal urban economic activity involves contact with new people on a regular basis, likely in HD environments (construction work, small firms etc.) that could increase costs of diversity. Eventually as workplaces become more formal with a regular set of colleagues, these costs might reduce again. One important aspect of this is that identity diversity might act as a hindrance in transitioning to formal work by perpetuating discrimination amongst groups (Awaworyi Churchill and Danquah, 2020). As a result, policies that promote trust between ethnic groups including contact that shapes positive preferences will be important.

Finally, the finding that minorities (Muslims) bear the cost of integration in this context is generalizable to many other settings, especially the U.S. For example, the argument that African-Americans are rewarded less for their effort (relative to the average American), requiring them to work harder to achieve similar career goals (DeSante, 2013) or the finding that Asian immigrants in the U.S., being aware of their unequal racial status, *work twice as hard* as a normative path to success and assimilation by achieving model-minority status (Zhou, 2004 and Zhou and Xiong, 2005), relate closely to my results in the Indian context. Overall, this implies that minority (the oppressed) groups, despite being discriminated against, are likely to play a crucial in the process of nation-building through initiating economic and social integration in diverse societies. But the above statement must be caveated with the fact that we still have much left to understand with respect to if/how this would also translate into assimilation in the sense that minorities do not feel a divide between participation in mainstream institutions and religious and cultural practices.

References

Afridi, F., A. Dhillon, S. X. Li, and S. Sharma (2020). Using social connections and financial incentives to solve coordination failure: A quasi-field experiment in India's manufacturing sector. *Journal of Development Economics 144*, 102445.

Afridi, F., A. Dhillon, and S. Sharma (2020). The ties that bind us: Social networks and productivity in the factory. *CEPR Discussion Paper No. DP14687*.

Alam, M. S. (2010). Social exclusion of Muslims in India and deficient debates about affirmative action: Suggestions for a new approach. *South Asia Research 30*(1), 43–65.

Alesina, A. and E. Spolaore (1997). On the number and size of nations. *The Quarterly Journal of Economics 112*(4), 1027–1056.

Allport, G. W., K. Clark, and T. Pettigrew (1954). The nature of prejudice. *Addison-wesley Reading, MA*.

Altonji, J. G. and C. R. Pierret (1998). Employer learning and the signalling value of education. pp. 159–195.

Altonji, J. G. and C. R. Pierret (2001). Employer learning and statistical discrimination. *The Quarterly Journal of Economics 116*(1), 313–350.

Artiles, M. (2020). Within-group heterogeneity in a multi-ethnic society. *Working Paper*.

Asher, S., P. Novosad, and C. Rafkin (2018). Intergenerational mobility in india: Estimates from new methods and administrative data. *World Bank Working Paper*.

Ashraf, N. and O. Bandiera (2018). Social incentives in organizations. *Annual Review of Economics 10*, 439–463.

Awaworyi Churchill, S. and M. Danquah (2020). Ethnic diversity and informal work in Ghana. Technical report, WIDER Working Paper.

Bandiera, O., I. Barankay, and I. Rasul (2010). Social incentives in the workplace. *The Review of Economic Studies 77*(2), 417–458.

Bandiera, O., I. Barankay, and I. Rasul (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association 11*(5), 1079–1114.

Basant, R. (2007). Social, economic and educational conditions of Indian Muslims. *Economic and Political Weekly*, 828–832.

Bazzi, S., A. Gaduh, A. D. Rothenberg, M. Wong, et al. (2017). Unity in diversity?: Ethnicity, migration, and nation building in Indonesia. *Centre for Economic Policy Research*.

Becker, G. (1957). The economics of discrimination. *Chicago: University of Chicago Press*.

Bhalotra, S. R., I. Clots-Figueras, L. Iyer, and J. Vecci (2018). Leader identity and coordination. *IZA Discussion Paper*.

Bhaumik, S. K. and M. Chakrabarty (2009). Is education the panacea for economic deprivation of Muslims?: Evidence from wage earners in India, 1987–2005. *Journal of Asian Economics 20*(2), 137–149.

Broockman, D. and J. Kalla (2016). Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science 352*(6282), 220–224.

Burns, J., L. Corno, and E. La Ferrara (2015). Interaction, prejudice and performance: Evidence from South Africa. *Unpublished Working Paper*.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics 90*(3), 414–427.

Carpenter, J. and E. Seki (2011). Do social preferences increase productivity? Field experimental evidence from fishermen in Toyama Bay. *Economic Inquiry 49*(2), 612–630.

Carrell, S. E., M. Hoekstra, and J. E. West (2015). The impact of intergroup contact on racial attitudes and revealed preferences. *National Bureau of Economic Research*.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2019). On binscatter. *arXiv preprint arXiv:1902.09608*.

Churchill, S. A., M. R. Valenzuela, and W. Sablah (2017). Ethnic diversity and firm performance: Evidence from China's materials and industrial sectors. *Empirical Economics 53*(4), 1711–1731.

DeSante, C. D. (2013). Working twice as hard to get half as far: Race, work ethic, and America's deserving poor. *American Journal of Political Science 57*(2), 342–356.

Easterly, W. and R. Levine (1997). Africa's growth tragedy: Policies and ethnic divisions. *The Quarterly Journal of Economics 112*(4), 1203–1250.

Farber, H. S. and R. Gibbons (1996, 11). Learning and Wage Dynamics*. *The Quarterly Journal of Economics 111*(4), 1007–1047.

Fershtman, C. and U. Gneezy (2001). Discrimination in a segmented society: An experimental approach. *The Quarterly Journal of Economics 116*(1), 351–377.

Hamilton, B. H., J. A. Nickerson, and H. Owan (2012). Diversity and productivity in production teams. In *Advances in the Economic Analysis of participatory and Labor-managed Firms*. Emerald Group Publishing Limited.

Hjort, J. (2014). Ethnic divisions and production in firms. *The Quarterly Journal of Economics 129*(4), 1899–1946.

Hsieh, C.-T. and B. A. Olken (2014). The “missing middle". *Journal of Economic Perspectives 28*(3), 89–108.

Jakiela, P., E. Miguel, and V. te Velde (2011). Combining field and lab experiments to estimate the impact of human capital on social preferences. *Mimeo, UC Berkeley*.

Kalpagam, U. et al. (2010). Are Muslims discriminated against in the labour market in India? *Indian Journal of Labour Economics 53*(1).

Khan, J. I. (2019). Muslims in Indian labour market: Access and opportunities.

Kremer, M. (1993). The O-ring theory of economic development. *The Quarterly Journal of Economics 108*(3), 551–575.

Lazear, E. P. (1998). Personnel economics for managers. *Wiley New York*.

Lowe, M. (2020). Types of contact: A field experiment on collaborative and adversarial caste integration. *CESifo Working Paper*.

Marx, B., V. Pons, and T. Suri (2021). Diversity and team performance in a kenyan organizationâ. *Journal of Public Economics 197*, 104332.

Mas, A. and E. Moretti (2009). Peers at work. *American Economic Review 99*(1), 112–45.

Miguel, E. (2004). Tribe or nation? nation building and public goods in Kenya versus Tanzania. *World politics 56*(3), 327–362.

Mitra, A. and D. Ray (2014). Implications of an Economic Theory of Conflict: Hindu-Muslim Violence in India. *Journal of Political Economy 122*(4), 719–765.

Montalvo, J. G. and M. Reynal-Querol (2017). Ethnic diversity and growth: Revisiting the evidence. *Review of Economics and Statistics*, 1–43.

Mousa, S. (2018). Overcoming the trust deficit: Inter-group contact and associational life in post-ISIS Iraq. *Science. Vol. 369, Issue 6505, pp. 866-870.*.

Paluck, E. L., S. A. Green, and D. P. Green (2019). The contact hypothesis re-evaluated. *Behavioural Public Policy 3*(2), 129–158.

Parrotta, P., D. Pozzoli, and M. Pytlikova (2012). Does labor diversity affect firm productivity? *IZA Discussion Paper*.

Pettigrew, T. F., L. R. Tropp, U. Wagner, and O. Christ (2011). Recent advances in intergroup contact theory. *International journal of intercultural relations 35*(3), 271–280.

Rao, G. (March, 2019). Familiarity does not breed contempt: Generosity, discrimination and diversity in Delhi schools. *American Economic Review Vol. 109, No. 3.*

Schindler, D. and M. Westcott (2015). Shocking racial attitudes: The cultural legacy of black GIs in Europe. *Mimeo, University of Munich, Munich.*

Zhou, M. (2004). Are Asian-Americans becoming "White?". *Contexts 3*(1), 29–37.

Zhou, M. and Y. S. Xiong (2005). The multifaceted American experiences of the children of Asian immigrants: Lessons for segmented assimilation. *Ethnic and Racial Studies 28*(6), 1119–1152.

Appendix

A Randomization steps, implementation timeline and balance (Identification) checks

A.1 Randomization steps and timeline

Each step in the randomization process is described in detail below.

Step 0: Determine religious composition of each section in each line

For each section of each line, first decide final number of Hindus and Muslims (typically 40% Muslims in mixed sections)

s.t. $\sum Hs = \bar{H}$ *and* $\sum Ms = \bar{M}$, where $\bar{H}$ and $\bar{M}$ denote the total number of Hindus and Muslims in the line across all 3 shifts.

Workers were not moved across production lines for randomization. Therefore, the religious composition of line-section-level teams was constrained by the overall number of Hindus and Muslims in the line at baseline. However, since the religious composition of each line was very similar to the overall composition in the factory, it meant there would be 40% Muslims in mixed sections (HD or LD).

Step 1: Section Shifting

Suppose 2 additional Muslim workers are required in a section to achieve the desired religious composition (40% Muslims). Then the following steps are taken:

a) Randomly order within section × religion × skill

b) Find a section with enough Muslims

c) Randomly pick 2 Muslim workers to shift in

d) Randomly pick 2 Hindu workers to shift out

This step is perhaps the most crucial in order to achieve the desired line-level treatment types described in Figure 3. Since at baseline, not all sections of all lines (across all 3 cohorts) had enough Muslim workers to have 40% Muslims in mixed line-section-level teams post randomization, workers were moved across sections in this manner to achieve that. This also meant, only the minimum number of workers required were moved, satisfying the firm's requirement of minimizing section-switching.

Step 2: Re-randomize

a) Randomly order within new section × religion × skill level

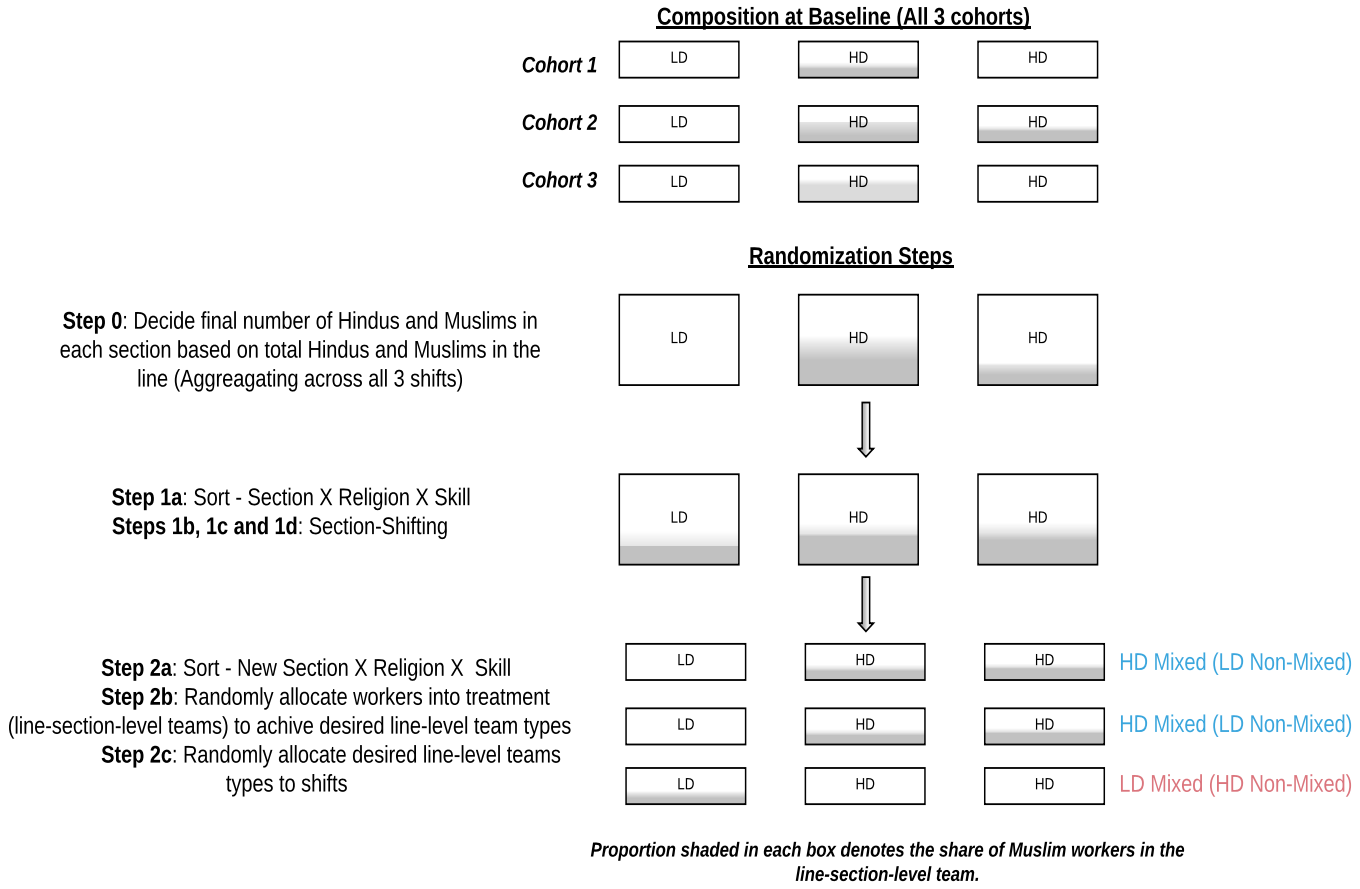
b) Allocate workers into mixed vs homogeneous teams as pre-specified

c) Randomly allocate teams (lines) to shifts/supervisors

In Step 2, workers were sorted by their new section (which is largely the same as the old section for most workers, but different for those that were moved in Step 1), religion and skill and allocated to line-section-level teams (recall that there are three teams per section in a line, one for each shift). The line-section-level teams were then aggregated and line-level teams were formed to achieve the treatment types, as in Figure 3. Finally, these line-level teams as whole were randomly allocated to the three shifts and the usual weekly shift rotations were introduced.

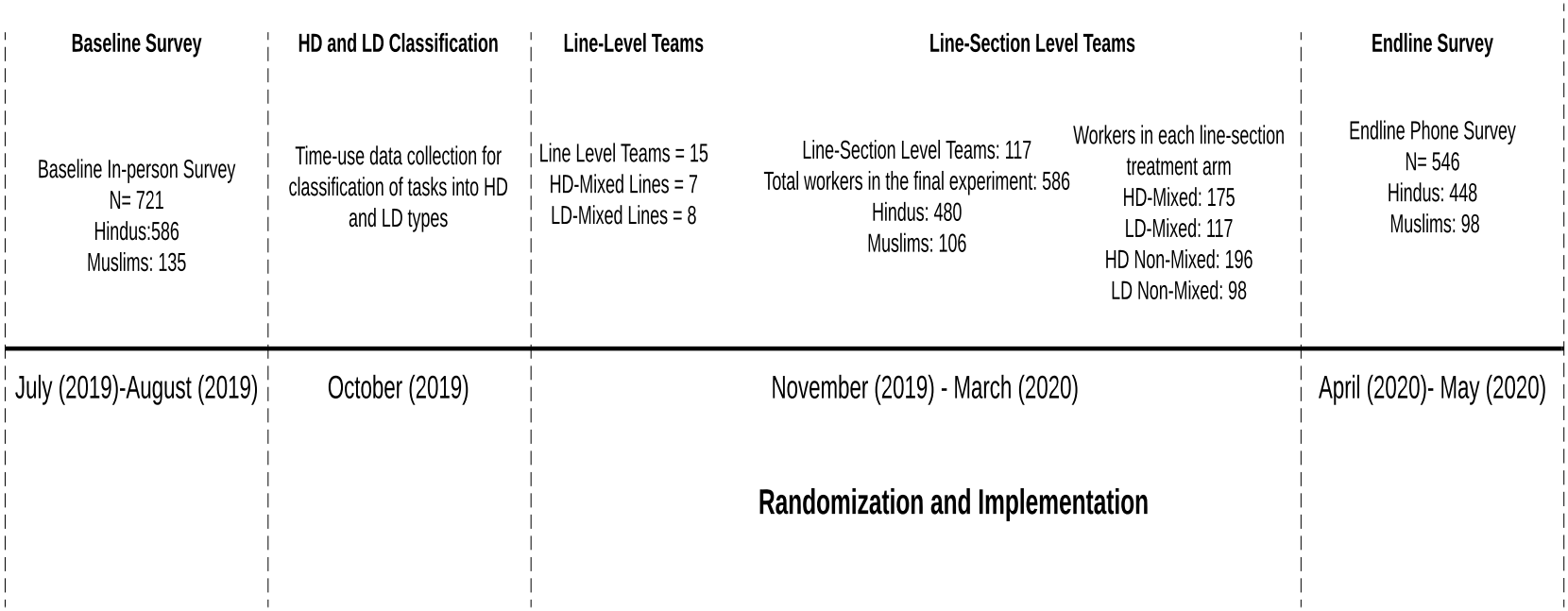
Figure A.1 illustrates how the desired line-level team types are achieved given the religious composition of sections of a production in all 3 shifts at baseline. It essentially provides a visual representation of the steps described above.

Figure A.1: Randomized Steps



Note: This figure illustrates the steps involved in the randomization process – from how given the religious composition of sections at baseline the desired line-level team types are achieved. The figure is based on the description of the steps discussed in section A.1.

Figure A.2: Experiment Design and Timeline



Note: This diagram shows the timeline of the intervention. The baseline survey was completed between July and August in 2019. Time-use data in order to classify tasks into HD and LD types was collected in October 2019. The randomization and the actual experiment was conducted between November 2019 and March 2020. Detailed sample size by treatment arms is presented in the figure. The baseline survey included more workers than in the actual experiment, because the factory laid off workers before the start of the intervention due to low product demand in one production line. A endline phone survey was conducted in April and May of 2020 due to COVID-19 related restrictions.

A.2 Checks for dependency sorting

Hiring at the factory occurs on a rolling basis as and when vacancies become available for each position on a production line. The HR manager always has a pool of job applicants at hand who are called upon on a first-come-first-serve basis. As a result, workers do not have the option to choose their area of work when they join. It is possible that workers quit at a different rate across the two types of sections, leading to possible selection bias. However, if that were the case, this would be reflected in the average tenure of workers in HD and LD sections. As shown in Table A.4, this is not the case - tenure is balanced between workers in HD and LD sections.

While selection into jobs is therefore unlikely at hiring, it is possible that over time, workers are able to sort into their sections of choice. In order to assess if that is the case workers were asked to report their first job at the factory and their final job immediately before the intervention began. They were also asked to report any other job that they held for a period of more than 6 months at the factory. Table A.1 reports a matrix of job switches between HD and LD sections. Only 94 out of 586 workers (16%) reported to be currently in jobs that involved switching Dependency from their first job. Only 7.2% of workers overall reported to switch jobs more than once, whereby majority of workers who switched jobs did so only once. Additionally, many of these changes resulted from a closure of one production line at the factory in 2018 as a result of which workers in that line were reallocated, typically to similar jobs, in the same shift, but to other existing lines and an additional line which was bought around the same time.

Overall, this suggests that only a small share of workers switched jobs from when they first joined, until the time of the intervention. This rules out systematic sorting into tasks over time and possible selection bias resulting from it.

In Table A.2, I test whether observable characteristics of workers are correlated with the probability of moving across task types, based on the moves that have occurred, as shown in Table A.1. As it can be observed, none of the factors (age, tenure, religion) which could potentially affect sorting over time, are significant in Column (1). In Columns (2) and (3), I split up

job switches from HD to LD and LD to HD sections. Again the usual factors remain small in magnitude and not statistically significant.

Table A.1: Dependency Switches

First Job/Final Job	Low Dependency	High Dependency	Total
Low Dependency	148	35	183
High Dependency	59	344	403
Total	207	379	586

Note: This matrix reports the number of workers who, from when they first joined the factory until before the intervention, switched jobs that also involved switching dependencies. 35 workers (5.9%) switched from Low to High Dependency, while 59 workers (10%) switched from High to Low Dependency. Of the 15.9% workers who switched jobs, only 6.85% held one or more job between their first and final job at the factory.

In Column (2) however, it can be seen that workers who are currently Operators are likely to have switched from HD to LD at a higher rate than workers of other skill levels. Table A.3 shows that this difference is actually a result of the one-time move of workers from the production line shut in 2018 to other lines. If I leave this set of workers out of the analysis (as in Table A.3), Operators are no more likely to have switched from HD to LD than other workers. This is understandable since Operators in the line that has now been shut were all in Packing sections, which was a HD section in that line. However, Packing sections in the 6 production lines that are a part of this experiment are a combination of both HD and LD types. As a result, some Operators mechanically moved from HD to LD jobs when this change occurred, despite continuing to be Packing Operators in terms of their specific role in the production line.

In Table A.4, I report balance in work characteristics across treatment arms without the inclusion of line × section fixed effects. Therefore, unlike in Table 4, the main effect of being in HD versus LD section is identified. It can be observed that worker characteristics are balanced between HD and LD sections.

Table A.2: Dependency Sorting

	(1) **Switched Dependency**	(2) **High to Low**	(3) **Low to High**
Age	0.00353	-0.00264	-0.000890
	(0.00199)	(0.00183)	(0.00100)
Tenure	0.00190	-0.00405	0.00215
	(0.00664)	(0.00479)	(0.00378)
Muslim	-0.0144	0.0205	-0.00617
	(0.0388)	(0.0264)	(0.0273)
Worker Skill			
Semi-Skilled	-0.194	0.284	-0.0904
	(0.207)	(0.174)	(0.0511)
Operator	-0.0412	0.0942**	-0.0530
	(0.0536)	(0.0373)	(0.0312)
Line × Section F.E. (First Job)	Yes	Yes	Yes
N	579	579	579
Adj. R^2	0.067	0.094	0.284

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the worker's first line-section level. Workers were asked to report their first job at the factory and their last job before the intervention began. Switching refers to whether the move between the first and last job (if any) involved changing dependencies as well. Workers were also asked to report if they held any other job in between. Only 7.2 % reported that they did. Workers are categorized into the the following categories: unskilled, semi-skilled or operators. Unskilled workers are the omitted group in the regressions.

Table A.3: Dependency Sorting: Omitting workers shifted from shut production line

	(1) **Switched Dependency**	(2) **High to Low**	(3) **Low to High**
Age	0.00465*	-0.00334	-0.00132
	(0.00222)	(0.00211)	(0.00107)
Tenure	-0.00908	0.00348	0.00561
	(0.00586)	(0.00428)	(0.00432)
Muslim	-0.0290	0.00619	0.0228
	(0.0350)	(0.0263)	(0.0227)
Worker Skill			
Semi-Skilled	0.0512	0.0444	-0.0956
	(0.0822)	(0.0656)	(0.0539)
Operator	0.0117	0.0560	-0.0677
	(0.0619)	(0.0423)	(0.0388)
Line × Section F.E. (First Job)	Yes	Yes	Yes
N	470	470	470
Adj. R^2	0.045	0.002	0.267

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the worker's first line-section level. Workers were asked to report their first job at the factory and their last job before the intervention began. Switching refers to whether the move between the first and last job (if any) involved changing dependencies as well. Workers were also asked to report if they held any other job in between. Only 7.2 % reported that they did. Workers are categorized into the the following categories: unskilled, semi-skilled or operators. Unskilled workers are the omitted group in the regressions.

Table A.4: Randomization Check

	Panel A: Outcomes relevant at work				**Panel B: Other outcomes**				
	Tenure	Muslim co-workers	Orders	Communication	Age	Schooling	Trust	Altruism	Interact-cross
		Hindus							*Outside Work*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Mixed	0.0375	0.00893	0.0450	0.0707	1.305	-0.0307	0.591	0.0775	0.0382
	(0.464)	(0.0194)	(0.0586)	(0.0691)	(1.491)	(0.638)	(0.383)	(0.246)	(0.0516)
High Dependency	-0.584	0.0168	-0.00203	0.00637	1.128	-0.486	0.251	0.155	0.0134
	(0.400)	(0.0152)	(0.0624)	(0.0655)	(1.174)	(0.494)	(0.378)	(0.217)	(0.0498)
Mixed × High Dependency	-0.0315	0.00159	-0.0461	-0.118	-0.471	0.448	-0.640	-0.0913	-0.0239
	(0.564)	(0.0266)	(0.0872)	(0.0893)	(1.698)	(0.720)	(0.485)	(0.281)	(0.0650)
Mean Dep Var.	4.45	0.12	0.73	0.53	34.47	7.84	3.79	6.65	0.452
Line Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Religion Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	586	478	586	586	586	586	586	586	586
Adj. R^2	0.086	0.052	0.011	0.025	0.019	0.016	0.003	0.003	0.109

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the section-team level. Survey questions on “Trust" and “Altruism" are used from the World Value Survey (WVS). The dependent variable "Interact-Cross" refers to the cross-religion interaction of workers at baseline, outside of work.

B Treatment effect on Standard Output and Output Gap

Supervisors keep a log of standard (expected) output against actual output produced, in each shift for each line. This measure is based on inputs used in the production process; negative deviations from the standard level of output imply lower productivity and higher raw material wastage. In Table B.5 (Column 2), I use percentage deviation of actual output from standard output as the outcome variable. The formula for "Output Gap" is $\frac{ActualOutput-StandardOutput}{StandardOutput} * 100$. The coefficient estimate shows that on average, line-level teams with religiously mixed HD sections, fall short of their expected target by a greater degree than lines with religiously mixed LD sections. This happens despite the fact that on average HD-Mixed lines receive lower targets (Column 1), though this difference is not statistically significant (i.e. the treatment effect on standard output is not significant). This suggests that the treatment effects in Table 5 is a result of under-performance of workers in HD-Mixed lines and is not simply a product of differential target setting across treatment groups.

In Figure B.3, I plot histograms of deviation from standard output for each team type. There are a couple of important things to be observed in this figure. First, LD-Mixed lines have lower variance in terms of deviation from standard output relative to HD-Mixed lines. Secondly, in HD-Mixed lines, negative deviations from standard output occur with greater frequency. At the same time however, large positive deviations from standard output are not completely unusual in HD-Mixed lines. This suggests that religious diversity caused greater uncertainty (in terms of achieving daily targets) in HD mixed sections relative to LD mixed ones. This also points to the fact that output in mixed teams might be more susceptible to idiosyncratic shocks (religious events, conflict etc.) in HD sections due to the tight-knit nature of intergroup contact. This is formally tested in Table B.6. I generate rolling standard deviation measures of the Output Gap variable and report that standard deviation in Output Gap is higher in HD-Mixed (LD Non-Mixed) lines. I show robustness to a range of window sizes in generating the rolling stan-

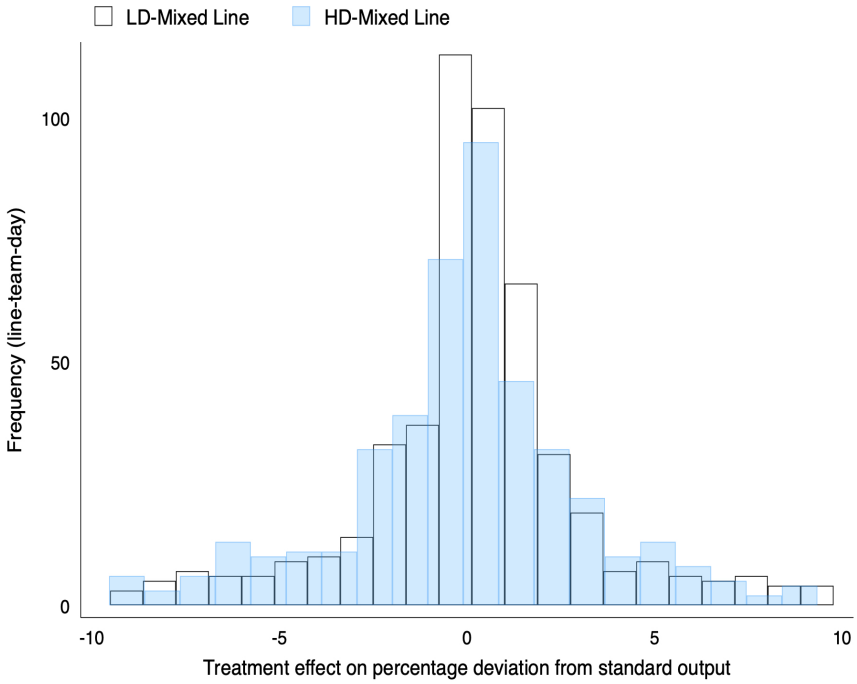
dard deviation measures. In Figure D.9, I plot cdfs of deviation of actual output from standard output and show that the probability actual output is greater than standard output is higher throughout, in LD-Mixed lines.

Table B.5: Treatment Effect on Line Level Standard Output

	(1)	(2)
	Log Standard Output	**Output Gap**
HD-Mixed Line (LD Non-Mixed)	-0.0223	-1.669***
	(0.0300)	(0.479)
Bootstrap Wild cluster C.I.	[-0.082, -0.0152]	[-3.833, -0.690]
Day Effects	Yes	Yes
Shift Effects	Yes	Yes
Production Line × Variety Effects	Yes	Yes
N	1045	1019
Adj. R^2	0.640	0.0488

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the production line (team) level in parenthesis. Wild cluster bootstrap confidence intervals in [] brackets. Standard Output is calculated from the amount of inputs used (batches mixed) in a shift and is predetermined. Output Gap gap is a measure of deviation from standard output and is calculated as $\frac{ActualOutput-StandardOutput}{StandardOutput} * 100$.

Figure B.3: Percentage deviation from standard output



Note: This figure shows percentage deviation from standard (expected) output by line-level treatment type. Each observation is a line-team on a particular day (shift) of production. The histograms that are not shaded denote LD-Mixed lines whereas the ones shaded in light blue denote HD-Mixed lines.

Table B.6: Treatment Effect on Standard Deviation of Output Gap

	Standard Deviation of Output Gap		
	(1)	(2)	(3)
Rolling Window	10 days	15 days	20 days
Min. Observations	5	10	15
HD Mixed Line (LD Non-Mixed)	0.954	2.784*	3.221**
	(0.857)	(1.358)	(1.342)
Bootstrap Wild cluster C.I.	[-1.38, 3.227]	[-1.394, 5.993]	[-1.583, 9.907]
Shift Effects	Yes	Yes	Yes
Production Line Effects	Yes	Yes	Yes
Day Effects	Yes	Yes	Yes
Mean Dep Var.	4.77	5.00	5.34
	(7.01)	(6.89)	(6.84)
N	755	404	231
Adj. R^2	0.374	0.552	0.668

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the production line-cohort-level in parenthesis. Wild cluster bootstrap (Cameron et al., 2008) confidence intervals in [] brackets. Rolling Window refers to the number of consecutive production days used to generate the standard deviation measure. Min. observations put a restriction on the minimum number of observations in each window.

C Additional tables

C.1 Summary Statistics

Table C.7 presents summary statistics of key aspects of the physical environment of HD and LD sections. Please refer to section 2 for a detailed description of this table.

Table C.7: Summary Statistics: Mean Differences (Physical Environment)

Variable	**Low Dependency**	**High Dependency**	**Diff (2) - (1)**
Panel A: Interaction (Minutes out of 10)			
Direct Dependency	2.22	9.50	7.283***
	(0.63)	(0.11)	(0.688)
Non-work Interaction	0.89	1.14	0.249
	(0.22)	(0.25)	(0.329)
Panel B: Noise Level (Decibels)			
Avg Noise (Db)	78.47	77.53	-0.941
	(1.42)	(1.66)	(2.170)
Max Noise (Db)	87.46	85.40	-2.055
	(1.72)	(1.65)	(2.394)
Panel C: Temperature (Celsius)			
Section Temperature (°C)	29.08	31.42	2.341*
	(0.92)	(0.72)	(1.197)
N	22	20	42

* p<0.10, ** p<0.05, *** p<0.010. In some cases, certain sections can have more than one task, with their degrees of dependency highly correlated. Sections are classified based on the average dependency minutes in each section.

Table C.8 presents summary statistics of characteristics of Hindu and Muslim workers described in section 2.

Table C.8: Summary Statistics: Hindu and Muslim workers

Variable	**Hindu**	**Muslim**	**Diff (2) - (1)**
Panel A: Dependency			
High Dependency (share of workers)	0.645	0.641	-0.005
	(0.22)	(0.48)	(0.052)
N	477	103	580
Panel B: Schooling and Tenure			
Schooling (Grade)	8.41	6.98	-1.437***
	(0.13)	(0.26)	(0.293)
Tenure	4.41	2.67	-1.740***
	(0.14)	(0.30)	(0.332)
N	584	134	718
Panel C: Cross-religion interaction and attitudes			
Cross-religion interaction (outside work)	0.39	0.72	0.329***
	(0.02)	(0.03)	(0.036)
Comfortable taking orders from non-coreligionists	0.71	0.70	-0.010
	(0.02)	(0.04)	(0.043)
Would live next door to non-coreligionists	0.60	0.89	-0.292***
	(0.02)	(0.02)	(0.034)
Equally comfortable communicating with non-coreligionists	0.51	0.72	0.211***
	(0.02)	(0.03)	(0.041)
N	586	135	721
Panel D: Political			
Supports National Registrar of Citizens (NRC)	0.35	0.22	-0.129***
	(0.02)	(0.03)	(0.043)
N	514	112	626

* p<0.10, ** p<0.05, *** p<0.010. Standard errors in parentheses. "Cross-religion interaction (outside work)" is a categorical variable coded 1, 0.5 and 0 if an individual reported to come in contact with greater than 5, between 1 and 5 and 0 non-coreligionists respectively in their daily life outside work. "Comfortable taking orders from non-coreligionists" is coded 1, 0.5 and 0 if an individual reports "Always comfortable", "Sometimes uncomfortable" and "Always uncomfortable" respectively. "Equally comfortable communicating with non-coreligionists" is coded in a similar manner.

C.2 Robustness checks and additional results

Table C.9: Treatment Effect on Standard Output (Line × Variety Effects)

	(1)	(2)
	Log Output (Pieces)	**Total Output (Boxes)**
HD-Mixed Line (LD Non-Mixed)	-0.0473***	-109.9***
	(0.0137)	(25.67)
Bootstrap Wild cluster C.I.	[-0.082, -0.0152]	[-195.6, -38.33]
Day Effects	Yes	Yes
Shift Effects	Yes	Yes
Production Line × Variety Effects	Yes	Yes
Mean Dep Var	10.80	1564
	(1.24)	(1450)
N	1045	1045
Adj. R^2	0.885	0.742

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the production line (team) level in parenthesis. Wild cluster bootstrap confidence intervals in [] brackets. Standard Output is calculated from the amount of inputs used (batches mixed) in a shift and is predetermined. Output Gap gap is a measure of deviation from standard output and is calculated as $\frac{ActualOutput-StandardOutput}{StandardOutput} * 100$.

Table C.10: Treatment Effect on Standard Output (Line × Day Effects)

	(1)	(2)
	Log Output (Pieces)	**Total Output (Boxes)**
HD-Mixed Line (LD Non-Mixed)	-0.0520**	-58.74**
	(0.0185)	(22.98)
Bootstrap Wild cluster C.I.	[-0.092, -0.005]	[-105, 0.1541]
Shift Effects	Yes	Yes
Production Line × Day Effects	Yes	Yes
Mean Dep. Var.	10.80	1564
	(1.24)	(1450)
N	1045	1045
Adj. R^2	0.900	0.911

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the production line (team) level in parenthesis. Wild cluster bootstrap confidence intervals in [] brackets. Standard Output is calculated from the amount of inputs used (batches mixed) in a shift and is predetermined. Output Gap gap is a measure of deviation from standard output and is calculated as $\frac{ActualOutput-StandardOutput}{StandardOutput} * 100$.

Table C.11: Treatment Effect on Section Ratings

	HD Sections		**LD Sections**	
	Rating (1)	**Rating > Median** (2)	**Rating** (3)	**Rating > Median** (4)
Mixed	-0.0496***	-0.0499***	-0.0005	-0.0024
	(0.0184)	(0.0113)	(0.0142)	(0.0124)
Mean Dep. Var.	3.85	3.85	3.80	3.80
	(0.68)	(0.68)	(0.64)	(0.64)
Education and Tenure Controls	Yes	Yes	Yes	Yes
Day Effects	Yes	Yes	Yes	Yes
Shift Effects	Yes	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes	Yes
N	3466	3466	3443	3443
Adj. R^2	0.609	0.385	0.595	0.324

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-level team. Education and tenure control for the mean of schooling and tenure of workers in the line-section-level team.

Table C.12: Treatment Effect on Section Ratings (Without controls collinear with religion)

	Rating (Raw)		**Rating > Median**	
	(1)	(2)	(3)	(4)
Mixed	-0.0239**		-0.0245***	
	(0.0114)		(0.00846)	
Mixed × LD		-0.00836		-0.00403
		(0.0150)		(0.0126)
Mixed × HD		-0.0394**		-0.0449***
		(0.0175)		(0.0113)
p(Mixed × HD = Mixed × LD)		0.184		0.017
Mean Dep. Var.	3.82	3.82	0.44	0.44
	(0.83)	(0.83)	(0.50)	(0.50)
Education and Tenure Controls	No	No	No	No
Day Effects	Yes	Yes	Yes	Yes
Shift Effects	Yes	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes	Yes
N	6909	6909	6909	6909
Adj. R^2	0.600	0.600	0.358	0.358

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-level team. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified in columns (2) and (4). Education and tenure control for the mean of schooling and tenure of workers in the line-section-level team.

Table C.13: Section ratings and characteristics

	Ratings (Raw)		
Tenure	0.00796	0.00834*	0.00875
	(0.00484)	(0.00489)	(0.00511)
Schooling		-0.00267	-0.00282
		(0.00323)	(0.00323)
Group Size			0.0173
			(0.0474)
Education and Tenure Controls	Yes	Yes	Yes
Day Effects	Yes	Yes	Yes
Shift Effects	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes
N	6909	6909	6909
Adj. R^2	0.600	0.600	0.600

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-level team. Education and tenure control for the mean of schooling and tenure of workers in the line-section-level team.

Table C.14: Section Ratings and Religious Violence

	Raw Ratings	
	(1)	(2)
Mixed × LD	-0.00674	
	(0.0144)	
Mixed × HD	-0.0349**	
	(0.0185)	
Mixed × LD × No violence		0.00233
		(0.0155)
Mixed × LD × Violence		-0.0964***
		(0.0342)
Mixed × HD × No violence		-0.0309*
		(0.0185)
Mixed × HD × Violence		-0.0766*
		(0.0408)
Mean Dep. Var.	3.82	3.82
	(0.83)	(0.83)
Education and Tenure Controls	Yes	Yes
Day Effects	Yes	Yes
Shift Effects	Yes	Yes
Line × Section Effects	Yes	Yes
N	6909	6909
Adj. R^2	0.600	0.600

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-level team. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified. Education and tenure control for the mean of schooling and tenure of workers in the line-section-level team. Between 13th-18th December 2019, immediately after the passing of the Citizenship Amendment Act (CAA) violent protests erupted (mainly by Muslims) in the district of West Bengal where the factory is located. Hindu-Muslim riots occurred in Delhi between 23rd-27th Feb 2020 during protests against the CAA as well. These days are coded as violent days in these regressions.

Table C.15: Teammate Choices: Only Hindus

	Identified teammate as Free Rider		**Blamed by teammate**		**Unwilling to give up Relief Time**	
	(1)	(2)	(3)	(4)	(5)	(6)
Mixed	0.0363***		0.0365**		0.0532	
	(0.0128)		(0.0151)		(0.0338)	
Mixed × LD		0.0267		0.0617***		0.0242
		(0.0237)		(0.0233)		(0.0506)
Mixed × HD		0.0388**		0.0302*		0.0615
		(0.0149)		(0.0176)		(0.0413)
p(Mixed × LD = Mixed × HD)		0.665		0.282		0.573
Mean. Dep. Var.	0.130	0.130	0.072	0.072	0.69	0.69
Worker skill F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	3248	3248	3220	3220	3313	3313
Adj. R^2	0.014	0.013	0.012	0.012	0.047	0.046

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-team level. Observations are at the section worker-teammate level i.e. there are (N-1) observations per worker, where N denotes the number of workers in the section. Workers were asked to choose teammates who they: (1) think have been a free-rider since teams changed (2) they have been blamed by during the experiment and (3) would give up their relief time for.

Table C.16: Teammate Choices: Decomposition (Mixed Teams)

	Identified teammate as Free Rider		**Blamed by teammate**		**Unwilling to give up Relief Time**	
	(1)	(2)	(3)	(4)	(5)	(6)
Target Muslim	0.0534***	0.101***	0.00159	0.0305	0.0341**	-0.0745
	(0.0160)	(0.0225)	(0.0187)	(0.0278)	(0.0160)	(0.0405)
Respondent Muslim	-0.0175	0.0286	-0.0134	0.0163	-0.0556*	-0.163***
	(0.0204)	(0.0258)	(0.0116)	(0.0208)	(0.0285)	(0.0382)
Target Muslim ×		-0.135***		-0.0840**		0.305***
Respondent Muslim		(0.0450)		(0.0388)		(0.0797)
Mean. Dep. Var.	0.146	0.146	0.091	0.091	0.331	0.331
Worker skill F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	2165	2165	2158	2158	2185	2185
Adj. R^2	0.018	0.025	0.013	0.016	0.064	0.084

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-team level. Observations are at the section worker-teammate level i.e. there are (N-1) observations per worker, where N denotes the number of workers in the section. Workers were asked to choose teammates who they: (1) think have been a free-rider since teams changed (2) they have been blamed by during the experiment and (3) would give up their relief time for.

Table C.17: Teammate Choices: Decomposition (Mixed Teams by Dependency)

	Identified teammate as Free Rider		**Blamed by teammate**		**Unwilling to give up Relief Time**	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: HD Sections						
Target Muslim	0.0704***	0.112***	-0.0113	0.0153	0.0452**	-0.0325
	(0.0174)	(0.0275)	(0.0238)	(0.0342)	(0.0189)	(0.0476)
Respondent Muslim	-0.0206	0.0176	-0.0108	0.0182	-0.0519	-0.125**
	(0.0259)	(0.0318)	(0.0135)	(0.0248)	(0.0372)	(0.0453)
Target Muslim ×		-0.118*		-0.0825*		0.220**
Respondent Muslim		(0.0571)		(0.0473)		(0.0945)
Mean Dep. Var.	0.152	0.152	0.086	0.086	0.350	0.350
N	1668	1668	1676	1676	1667	1667
Adj. R^2	0.022	0.026	0.016	0.020	0.078	0.087
Panel B: LD Sections						
Target Muslim	0.00704	0.0721*	0.0377	0.0826**	0.00476	-0.192***
	(0.0228)	(0.0310)	(0.0209)	(0.0379)	(0.0260)	(0.0550)
Respondent Muslim	-0.00777	0.0640	-0.0213	0.0188	-0.0632*	-0.284***
	(0.0294)	(0.0393)	(0.0230)	(0.0364)	(0.0338)	(0.0358)
Target Muslim ×		-0.181***		-0.110*		0.543***
Respondent Muslim		(0.0543)		(0.0628)		(0.121)
Mean Dep. Var.	0.118	0.118	0.108	0.108	0.272	0.272
N	497	497	482	482	518	518
Adj. R^2	0.003	0.019	0.002	0.007	-0.003	0.080

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-team level. Observations are at the section worker-teammate level i.e. there are (N-1) observations per worker, where N denotes the number of workers in the section. Workers were asked to choose teammates who they: (1) think have been a free-rider since teams changed (2) they have been blamed by during the experiment and (3) would give up their relief time for.

Table C.18: Heterogeneous Treatment Effects (HD Section Ratings)

	Raw Ratings		
	(1)	(2)	(2)
Panel A: Past Contact			
Mixed	-0.0496***	-0.119**	-0.0871**
	(0.0184)	(0.0505)	(0.0460)
Average Tenure (Hindus)		-0.0139	
		(0.00971)	
Mixed × Average Tenure (Hindus)		0.0162*	
		(0.00991)	
Share of Muslims in previous team (Hindus)			-0.147
			(0.292)
Mixed × Share of Muslims in previous team (Hindus)			0.306
			(0.338)
Panel B: Political (taste) preferences			
Mixed	-0.0496***	0.0602	0.0102
	(0.0184)	(0.0720)	(0.0309)
Share BJP voters (Hindus)		0.0136	
		(0.0620)	
Mixed × Share BJP voters (Hindus)		-0.218*	
		(0.118)	
Share supports NRC (Hindus)			0.127
			(0.0610)
Mixed × Share supports NRC (Hindus)			-0.173**
			(0.0733)
Mean Dep. Var.	3.85	3.85	3.85
	(0.68)	(0.68)	(0.68)
Education and Tenure controls	Yes	Yes	Yes
Day Effects	Yes	Yes	Yes
Shift Effects	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes
N	3466	3466	3466
Adj. R^2	0.609	0.609	0.609

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-level team. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. “Average Tenure (Hindus)" refers to the mean of tenure value of Hindu workers in the line-section-level team. “Share BJP voters (Hindus)" and “Share supports NRC (Hindus)" are defined similarly based on data collected at baseline. Education and tenure control for the mean of schooling and tenure of workers in the line-section-level team.

Table C.19: Heterogeneous attenuation by preferences at baseline (HD Section Ratings)

Sample	**Full**	**Contact at Baseline**		**Support for BJP and NRC**	
		Above Median	Below Median	Above Median	Below Median
	(1)	(2)	(3)	(4)	(5)
Mixed × 0-60 days	-0.0661*	-0.0458	-0.182***	-0.125***	-0.0496
	(0.0391)	(0.0498)	(0.0595)	(0.0438)	(0.0619)
Mixed × 61-120 days	-0.0355	-0.0195	-0.0802**	-0.0792***	0.0177
	(0.0234)	(0.0429)	(0.0337)	(0.0242)	(0.0712)
Mean Dep. Var.	3.85	3.83	3.89	3.87	3.83
	(0.68)	(0.69)	(0.67)	(0.67)	(0.70)
Education and Tenure controls	Yes	Yes	Yes	Yes	Yes
Day Effects	Yes	Yes	Yes	Yes	Yes
Shift Effects	Yes	Yes	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes	Yes	Yes
N	3466	1884	1582	2384	1082
Adj. R^2	0.609	0.633	0.596	0.602	0.631

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-level team. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. In column (2), the sample consists of all line-section-level teams in which average Muslims teammates that Hindus in that team had at baseline is above median. In column (3), the sample consists of all line-section-level teams in which average Muslims teammates that Hindus in that team had at baseline is below median. In column (4), the sample consists of all line-section-level teams with support for the BJP or the NRC (averaged across all Hindu workers in the team) above median. In Column (5), the sample consists of all line-section-level teams with support for the BJP or the NRC (averaged across all Hindu workers in the team) below median.

Table C.20: Heterogeneous Treatment Effect: Worker Skill (Hindus)

Comfortable:	**Attitudes towards Muslims**					
	Taking Orders		**Communicating**		**Co-working**	
	(1)	(2)	(3)	(4)	(5)	(6)
Mixed	0.123***	-0.00981	0.0913**	0.0454	0.110***	0.0576
	(0.0439)	(0.0925)	(0.0406)	(0.0726)	(0.0349)	(0.0614)
Unskilled		-0.105		-0.116*		-0.114**
		(0.0639)		(0.0590)		(0.0511)
Mixed × Unskilled		0.176*		0.0650		0.0732
		(0.0941)		(0.0791)		(0.0755)
(1) + (3)		0.165***		0.110**		0.131***
		(0.0450)		(0.0452)		(0.0421)
Baseline controls	Yes	Yes	Yes	Yes	Yes	Yes
Line × Section Effects	Yes	Yes	Yes	Yes	Yes	Yes
N	448	448	448	448	448	445
Adj. R^2	0.075	0.081	0.069	0.071	0.056	0.058

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-team level. “Orders" refers to whether a worker is comfortable listening to orders from a non-coreligionist. "Communication" refers to whether a worker is comfortable communicating with non-coreligionists outside of work. Co-working refers to whether a worker is comfortable being put in a team with a non-coreligionist in future.

Table C.21: Treatment Effect: Worker Skill

	Orders		**Communication**	
Treatment Group	Baseline Mean	Endline Mean	Baseline Mean	Endline Mean
Non-Mixed and Skilled	0.79	0.80	0.59	0.61
	(0.40)	(0.39)	(0.42)	(0.40)
Mixed and Skilled	0.76	0.79	0.57	0.63
	(0.43)	(0.42)	(0.40)	(0.39)
Non-mixed and Unskilled	0.73	0.70	0.44	0.48
	(0.43)	(0.45)	(0.43)	(0.41)
Mixed and Unskilled	0.71	0.84	0.47	0.59
	(0.46)	(0.35)	(0.42)	(0.42)

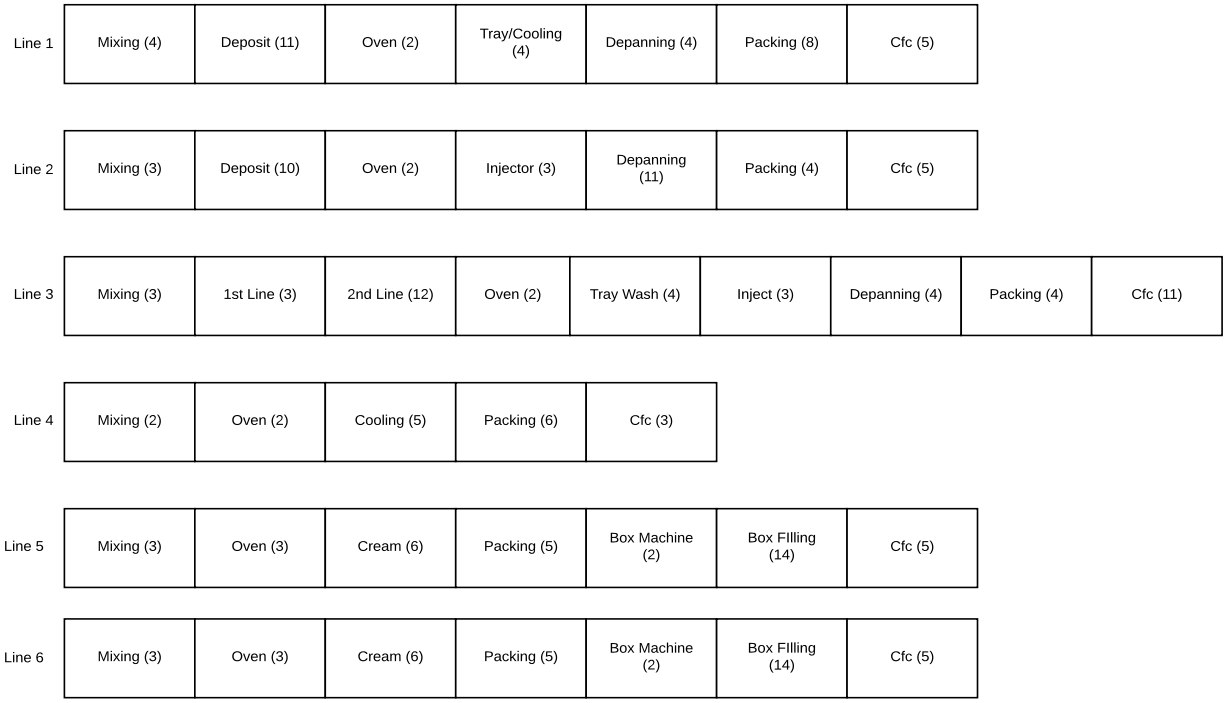
Table C.22: Section Change and Treatment Status

	Changed Section			
	(1)	(2)	(3)	(4)
Mixed	-0.0338	-0.0288		
	(0.0277)	(0.0249)		
Mixed × LD			-0.0353	0.0188
			(0.0538)	(0.0351)
Mixed × HD			-0.0329	-0.0551*
			(0.0361)	(0.0322)
Religion Effects	Yes	Yes	Yes	Yes
Line × Section F.E.	No	Yes	No	Yes
Line × Old Section F.E.	Yes	No	Yes	No
N	586	586	586	586
Adj. R^2	0.043	0.030	0.041	0.033

* p<0.10, ** p<0.05, *** p<0.010. Standard errors clustered at the line-section-team level. “Mixed" is a dummy variable coded 1 if the line-section-level team is religiously mixed. Line × Sections fixed effects are included in the all specifications; as a result the main effect of HD versus LD is not separately identified in columns (3) and (4). These are individual worker-level regressions. The outcome variable is a dummy coded 1 if after the randomization process the worker was in a different section (task) than their section of work at baseline.

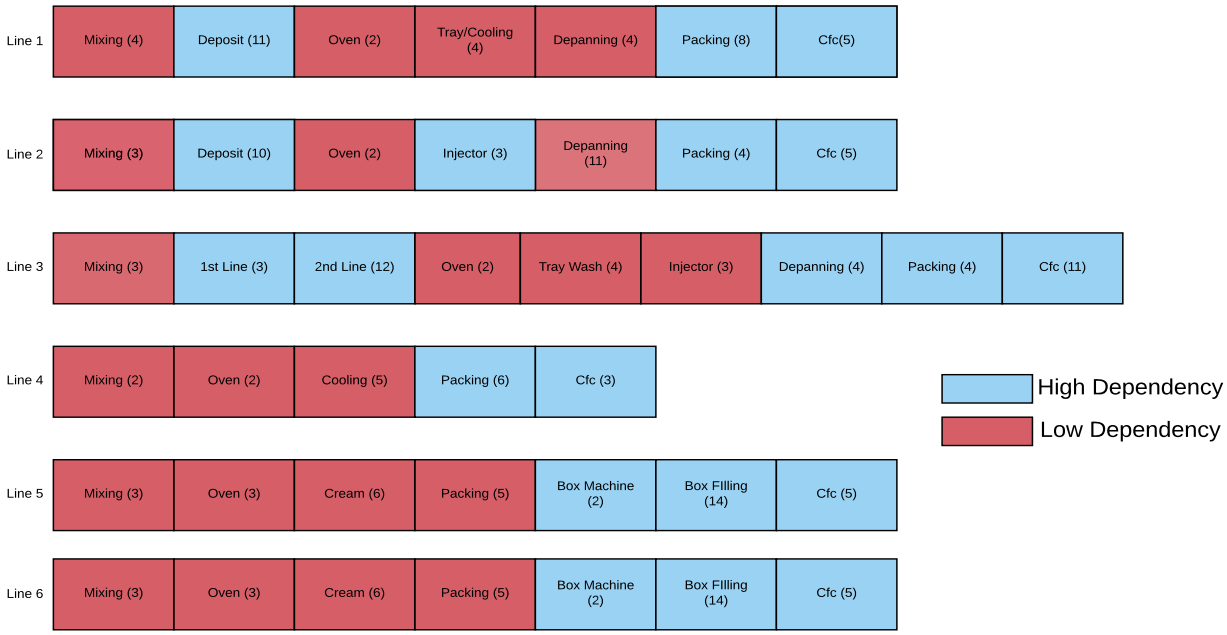
D Additional Figures

Figure D.4: Structure of Production Lines



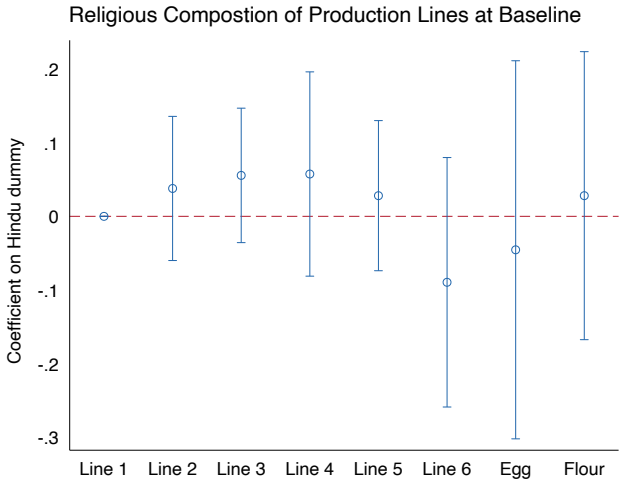
Note: This figure shows the structure of all 6 production lines in the factory. The numbers in parentheses denote the count of workers in each section per cohort. Each production line has 3 cohorts working on it in each of the 3 shifts in a day.

Figure D.5: High and Low Dependency Sections

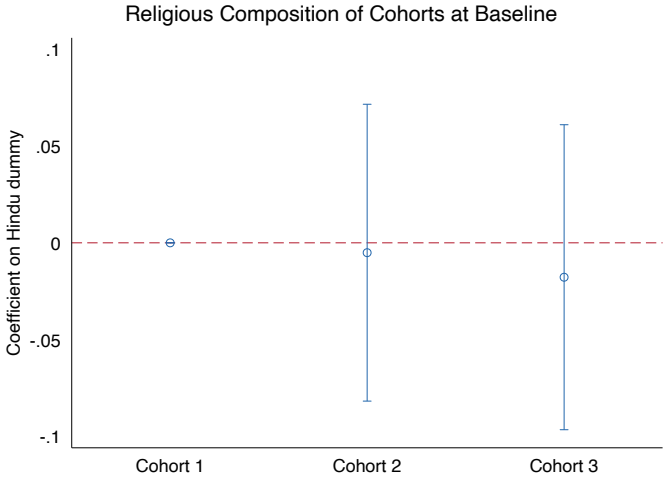


Note: This figure shows all sections of all lines at the factory split into HD and LD types. Direct Dependency is measures as described in section 2.2.

Figure D.6: Religious composition of lines and cohorts at baseline



(a) This figure plots coefficients from worker-level regressions. The outcome variable is a dummy coded 1 if the religion of the worker is Hindu and the independent variables are a set of dummy variables denoting each production line. “Egg" and “Flour" refer to production areas where raw materials (eggs and flour) are processed. These production areas are common to all production lines.



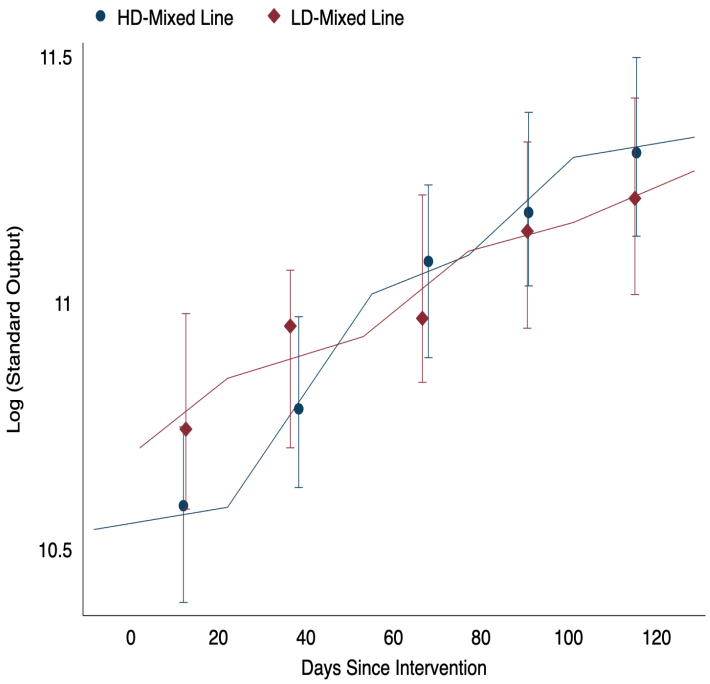
(b) This figure plots coefficients from worker-level regressions. The outcome variable is a dummy coded 1 if the religion of the worker is Hindu and the independent variables are dummies denoting cohorts (groups of workers who work at the factory at the same time).

Figure D.7: High and Low Dependency Tasks

(a) High Dependency

(b) Low Dependency

Figure D.8: Treatment Effect on Output (Event Study)



Note: This figure is generated from binned regressions using exactly the same controls variables as in Table 5. The treatment period is divided into 5 equal sized bins. The outcome variable is standard output (logged).

Figure D.9: Deviation from Standard Output

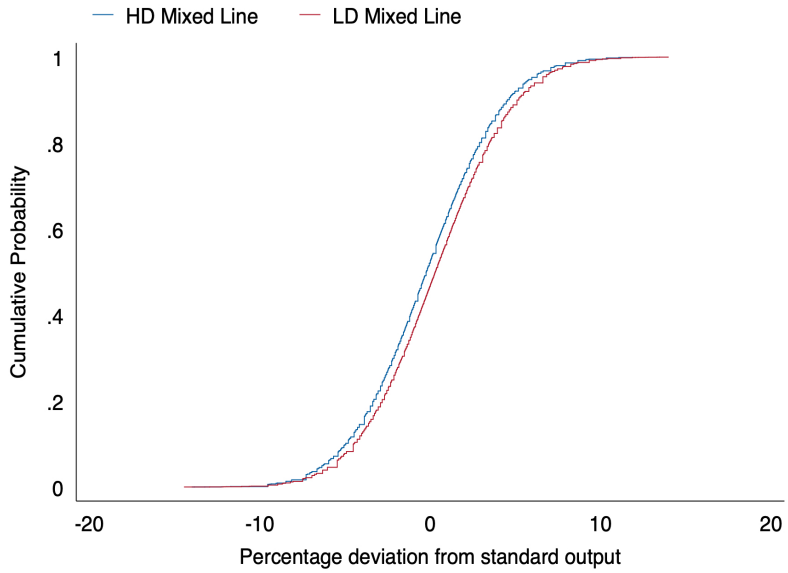
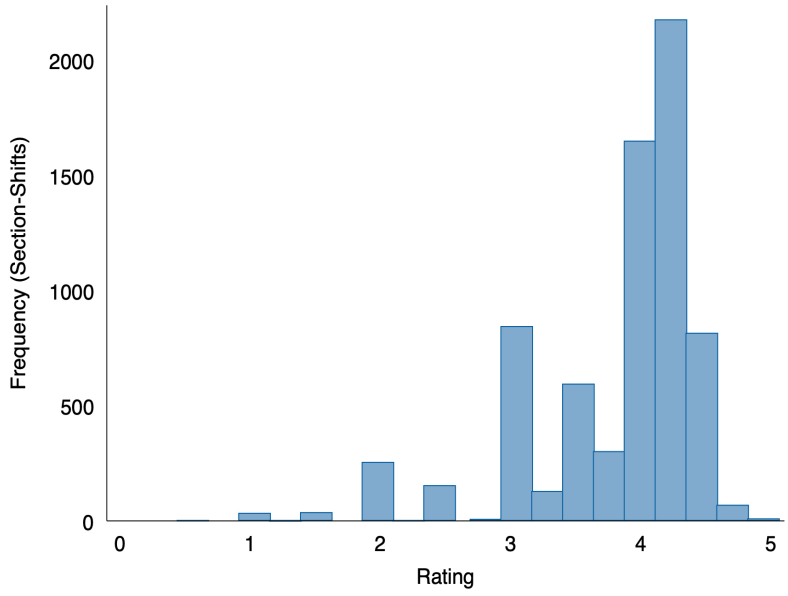


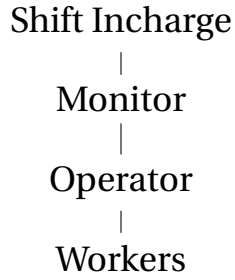
Figure D.10: Distribution of raw ratings



E Religion and Productivity: Firm Hierarchy

In this section, I provide additional evidence that Hindus and Muslims are not differentially productive, exploiting the hierarchical structure at the firm that used to exist up until early 2019, depicted in Figure E.11 below. This was previously the structure for each shift. At the very top, there were shift-incharges (one or two in a shift) who were the head of production during the shift and were responsible for overlooking the production process in multiple lines at the same time. Each line would then have a monitor responsible for maintaining performance of that line only and would report to their respective shift-incharge. Below that in the hierarchy was the packing operator. While there might be more than one operator in a line (based on the type of the product being made), the packing operator exists in all the lines. Finally came the set of workers working on the line.

Figure E.11: Hierarchical structure at the factory



While workers and operators did not change their area and shift of work, shift-incharges and monitors often did. This happened due to two reasons. Firstly and perhaps more importantly, the firm actively wanted incharges and monitors to gain experience of operation in all lines so that in the absence of one them (either because of absenteeism or if they quit) others can take up the position. This is important at this level in the hierarchy and not below because of the fewer number of workers in these positions and the greater complexity of work. Secondly, and related to the point above absenteeism (or being on leave) often meant that these workers

would switch their line and shift of work. Overall, between 2016-2018 I find that most incharges and supervisors had experience of working in all 6 lines at the factory. The representation of Muslims at this level in the hierarchy was close to their share in the factory overall (in fact a little greater than that at 22%). In Table E.23, I regress log of daily output at the line-cohort level on dummy variables denoting the religion of the personnel at each of these positions in a line-cohort.

Table E.23: Hindus and Muslims are equally productive (2016-2018)

	(1) **Log (Output)**
Muslim Supervisor	-0.0031 (0.0104)
Muslim Monitor	-0.0007 (0.0135)
Muslim Operator	0.0275 (0.0192)
Line Effects	Yes
Shift Effects	Yes
Month of year Effects	Yes
N	5727
Adj. R^2	0.848

* p<0.10, ** p<0.05, *** p<0.010. Robust standard errors in parenthesis.

The coefficients on the religion dummies for all three positions are small and not statistically significant suggesting that Hindus and Muslims are not differentially productive at the hierarchical level. These results complement the analysis from the experiment which shows no difference in output between LD Mixed and LD Non-Mixed teams, which I use as evidence that Hindus and Muslims are not differentially productive at the bottom of the hierarchical level

(amongst workers) as shown in Figure E.11.

F Model Appendix (Proofs)

Implication 1: *In a one shot game, the only equilibrium is where both workers exert low effort in a HD section.*

Proof: This follows from assumption (3) in the main text, which is:

$$p_H O^H + (1-p^H)O^L - c(e^H) > p_L O^L + (1-p^L)O^L > p_{HL} O^L + (1-p^{HL})O^L - c(eH) \tag{21}$$

Given that a Muslim worker expects a Hindu workers to initially exert low effort, it is optimal for the Muslim worker to also exert low effort given a short time horizon.

Implication 2: *Given a long enough horizon (large T), the optimal effort investment path for a Muslim worker would be to exert e_H from the beginning itself. Given that T is large enough, Hindu workers would eventually update $p > \bar{p}$ leading to higher expected steady state output (Bayesian Persuasion).*

Proof: Suppose a Hindu and a Muslim worker are working together in a team for periods t = $1, ..., T$, where T is finite but can be arbitrarily large. Since p < $\bar{p}$ at the beginning, the Hindu worker exerts low effort e_L. At time period 0, the Muslim worker maximizes future payoff which determines their optimal effort investment path. High effort is optimal in the beginning for the Muslim worker iff

$$V_k^{e_H} = \{ \sum_{s' \in S_{k+1}} P_k(s, e_H) + p_k(s'|s,a) V_{k+1}^{e_H}(s') \}_{a=e_H} \geq TP^{e_L} \tag{22}$$

for $k = N-1, ..., 0$. Suppose at time 0, (21) is true. Then, any other investment path rather than e_H at the beginning (specifically one where the Muslim worker initially exerts e_L and then e_H) is sub-optimal. To see this, suppose that this were not the case by contradiction. Then there

exists some t_1 and t_2 such that,

$$\begin{aligned} & t_1 P^{e_L}_{a(H)=e_L} + t_2 P^{e_H}_{a(H)=e_L} + (T - t_1 - t_2) P^{e_H}_{a(H)=e_H} \\ & \geq \{ \sum_{s' \in S_{k+1}} P_k(s, e_H) + p_k(s'|s,a) V^{e_H}_{k+1}(s') \}_{a=e_H, k=T-1,\ldots,0} \\ & \geq T P^{e_L} \end{aligned} \tag{23}$$

where t_1 and t_2 respectively denote time periods during which the Muslim worker puts low effort and high effort respectively (while the Hindu worker still has not updated their prior above $\bar{p}$). We can similarly split the payoff from exerting high effort and write the inequality as

$$\begin{aligned} t_1 P^{e_L}_{a(H)=e_L} + t_2 P^{e_H}_{a(H)=e_L} + (T - t_1 - t_2) P^{e_H}_{a(H)=e_H} & \geq \tilde{t_1} P^{e_H}_{a(H)=e_L} + (T - \tilde{t_1}) P^{e_H}_{a(H)=e_H} \\ & \geq T P^{e_L} \end{aligned} \tag{24}$$

Re-writing the above we have (and ignoring the last inequality),

$$\begin{aligned} & t_1 (P^{e_L}_{a(H)=e_L} - P^{e_H}_{a(H)=e_H}) + t_2 (P^{e_H}_{a(H)=e_L} - P^{e_H}_{a(H)=e_H}) + T P^{e_H}_{a(H)=e_H} \\ & \geq \tilde{t_1} (P^{e_H}_{a(H)=e_L} - P^{e_H}_{a(H)=e_H}) + T P^{e_H}_{a(H)=e_H} \end{aligned} \tag{25}$$

Notice that t_2 in expectation (at t= 0) must be larger than $\tilde{t_1}$. This is because in expectation the prior about the Muslim worker being a good type will be lower (than initial p) after T_1 periods of low effort by the Muslim worker i.e. $E(p_{T_1} | a(M) = e_L \in 0, \ldots, t_1; p) < p$. Hence the L.H.S in (14) cannot be greater than the R.H.S.

Now we prove (11). We split the payoff of the Muslim worker into two parts: before and after some arbitrary period j, such that for all t $\leq j$, $p_t \leq p$ and $p_t > p$ for $t > j$. Hence, we must show that:

$$t_j P^{e_H}_{a(H)=e_L} + (T - t_j) P^{e_H}_{a(H)=e_H} \geq T P^{e_L} \tag{26}$$

This involves showing there exists some $j <$ T, such that E[$p|a(M) = e_H$, for t = 1,....,j] > $\bar{p}$. Consider the following extreme scenarios and the consequent prior of the Hindu workers: (1) in each period before j, high output is produced and (2) in each period before j, low output is

produced. The priors in cases (1) and (2) respectively are:

$$(1) : \frac{p.p_{HL}^{j-1}}{p.p_{HL}^{j-1} + (1-p)p_L^{j-1}} \tag{27}$$

$$(2) : \frac{p.(1-p_{HL})^{j-1}}{p.(1-p_{HL})^{j-1} + (1-p)(1-p_L)^{j-1}} \tag{28}$$

(1) and (2) give the lower and upper bound on the beliefs of the Hindu worker about the type of the Muslim worker at time j. Any prior at time period j can therefore be written as a linear combination of expressions above. Re-writing, we therefore have,

$$\frac{p.p_{HL}^{j-1}}{p.p_{HL}^{j-1} + (1-p)p_L^{j-1}} + A.\frac{p.(1-p_{HL})^{j-1}}{p.(1-p_{HL})^{j-1} + (1-p)(1-p_L)^{j-1}} \tag{29}$$

$$= \frac{p}{p + (1-p)(\frac{p_L}{p_{HL}})^{j-1}} + A\frac{p}{p + (1-p)(\frac{1-p_L}{1-p_{HL}})^{j-1}} \tag{30}$$

where A is a negative constant. Now, as j gets large, the first term in the equation $\rightarrow 1$, while the second term $\rightarrow 0$. Therefore, provided T is large enough, the higher steady state output will compensate for the initial loss in payoff for the Muslim worker. This completes the proof.

Implication 3: *Convergence to high steady state output will be quicker the closer p is to $\bar{p}$ in the initial stage.*

Proof: This is trivial given the proof of Implication (2). Conditional on T, the closer initial p is to $\bar{p}$, fewer periods of high effort are required for the transition to expected high steady state output. The implication of this is that mixed worker teams with Hindus with a average higher p, should experience larger output relative to mixed teams in which Hindu workers have lower p.

Implication 4: *In LD sections, if Hindu and Muslim workers are not differentially productive, there will be no difference in output in mixed versus homogenous teams.*

Proof: Under the assumption that Hindu and Muslim workers are equally productive, the overall output of a LD section is simply the expected sum of individual outputs as shown below.

$$y_{LD}(e_i, ..., e_{LD}) = \sum_{i=1}^{I_{LD}} \{p_{ei} o_h + (1 - p_{ei}) o_l\} \tag{31}$$

We have also assumed that net expected payoff from high effort is greater than next expected payoff from low effort. In other words, in LD sections it is optimal for each worker to exert high effort. Hence on aggregate there should be no difference in output of mixed and non-mixed teams.